The relative effects of group size on reading progress of older students with reading difficulties

Sharon Vaughn · Jeanne Wanzek · Jade Wexler · Amy Barth · Paul T. Cirino · Jack Fletcher · Melissa Romain · Carolyn A. Denton · Greg Roberts · David Francis

Abstract This study reports findings on the relative effects from a yearlong secondary intervention contrasting large-group, small-group, and school-provided interventions emphasizing word study, vocabulary development, fluency, and comprehension with seventh- and eighth-graders with reading difficulties. Findings

S. Vaughn (✉) · J. Wexler · G. Roberts
The University of Texas at Austin, 1 University Station, D4900, Austin, TX 78712-0365, USA
e-mail: SRVaughnUM@aol.com

J. Wexler
e-mail: jwexler@mail.utexas.edu

G. Roberts
e-mail: gregroberts@mail.utexas.edu

J. Wanzek
Florida State University, P.O. Box 306-4304, Tallahassee, FL 32306, USA
e-mail: jwanzek@aol.com

A. Barth · P. T. Cirino
Texas Medical Center Annex, University of Houston, 2151 W. Holcombe Blvd., Suite 222, Houston, TX 77204-5053, USA

A. Barth
e-mail: amy.barth@times.uh.edu

P. T. Cirino
e-mail: pcirino@uh.edu

J. Fletcher · M. Romain
Texas Medical Center Annex, University of Houston, 2151 W. Holcombe Blvd., Suite 226, Houston, TX 77204-5053, USA

J. Fletcher
e-mail: jmfletch@Central.UH.EDU

M. Romain
e-mail: maronmain@Central.UH.EDU

indicate that few statistically significant results or clinically significant gains were associated with group size or intervention. Findings also indicate that a significant acceleration of reading outcomes for seventh- and eighth-graders from high-poverty schools is unlikely to result from a 50 min daily class. Instead, the findings indicate, achieving this outcome will require more comprehensive models including more extensive intervention (e.g., more time, even smaller groups), interventions that are longer in duration (multiple years), and interventions that vary in emphasis based on specific students' needs (e.g., increased focus on comprehension or word study).

Introduction

While considerable emphasis has been placed on learning to read during the past few decades, there has been increasing recognition that many older students demonstrate reading difficulties that significantly affect their reading to learn and reading for pleasure (Biancarosa & Snow, 2004; Kamil et al., 2008; Snow, 2002). These consensus reports document the relatively high numbers of older students (about one in four) who do not read and understand text at even a basic level and provide guidance for ongoing instruction and support for older students with reading difficulties. Although it is unlikely that these students will make accelerated progress without intensive interventions, there is evidence that secondary students may experience improved reading outcomes when provided explicit reading intervention with adequate time and intensity for reading instruction (Archer, Gleason, & Vachon, 2003; Torgesen et al., 2001).

Interventions for older students with reading difficulties

Though prevention of reading difficulties is ideal, many students have either untreated or persistent reading difficulties that continue into the secondary grades (Biancarosa & Snow, 2004). Interestingly, despite the prevailing perception that older students have low interest and motivation to read, qualitative reports based on repeated interviews with students (e.g., McCray, Vaughn, & La Vonne, 2001) reveal that they are embarrassed by their low reading skills and would be highly interested in learning to read if the reading intervention "actually worked."

———————————————

C. A. Denton
The University of Texas Health Science Center Houston, 7000 Fannin, UCT Suite 2443, Houston, TX 77030, USA
e-mail: Carolyn.A.Denton@uth.tmc.edu

D. Francis
Texas Institute for Measurement, Evaluation, and Statistics, University of Houston, 100 TLCC Annex, Houston, TX 77204-6022, USA
e-mail: dfrancis@uh.edu

Several recent syntheses provide useful summaries of intervention studies and effective practices for secondary readers with reading difficulties (Edmonds et al., 2009; Scammacca et al., 2007). In a meta-analysis of reading intervention studies for older students (grades 6–12) with reading difficulties, Edmonds et al. (2009) reported a mean weighted average effect size for reading comprehension of 0.89 favoring treatment over comparison students. Interventions that focused primarily on decoding were associated with moderate effect size gains in reading comprehension (ES = 0.49). Scammacca et al. (2007) reported a mean effect size of 0.95 for reading comprehension. Moderator analyses revealed that researcher-developed instruments were associated with larger effect sizes than standardized, norm-referenced measures (ES = 0.42). In addition, word study interventions were associated with moderate effects, with researcher-implemented interventions associated with higher effects than teacher-implemented interventions. Higher overall outcomes were associated with students in the middle grades rather than students in high school.

These two comprehensive syntheses indicate that older students with reading difficulties are likely to benefit from vocabulary, comprehension, and word study reading interventions. There are several issues to consider when interpreting the findings from these syntheses. First, the performance of the comparison groups influences effect sizes from these studies—as well as those from all intervention studies. If students in a comparison group are provided reading interventions, then the effects for the treatment group are likely to be lower than when comparison students are not provided reading intervention. Second, most of the interventions represented in the syntheses were less than 2 months in duration. In other studies, interventions provided relatively large effects initially (over the first 2 or 3 months), with reduced effects over time (Wanzek & Vaughn, 2007). Thus, short-term interventions may benefit from initial large impact. Third, insufficient data were available from the studies to determine whether the interventions improved student outcomes relative to grade-level expectations. With older students who are likely to be several grade levels behind peers, it is possible that even a large effect may not accelerate performance to a meaningful level relative to grade-level expectations.

Accelerating reading growth with older students

A significant goal of an intervention with older students is to accelerate student progress to reduce the gap between students' current reading level and grade-level expectations. Acceleration of this type requires a sufficiently intense intervention (Torgesen, Rashotte, Alexander, Alexander, & MacPhee, 2003), which for older readers requires consideration of adequate time on task with adequate opportunities to respond so that treatment can be appropriately situated to students' needs, maximizing students' outcomes (Heward & Silvestri, 2005). While there are several considerations for increasing intensity, there are two primary variables to consider: increased time for reading intervention and reduced group size so that students have targeted instruction with adequate feedback and support (Vaughn, Linan-Thompson, & Hickman, 2003). Neither of these variables has been manipulated in studies with older students with reading difficulties to determine their relative effects (Scammacca et al., 2007; Torgesen et al., 2007).

While increasing the amount of time each day that students are provided an intervention may be an acceptable strategy to researchers, it becomes challenging for secondary teachers and principals to find time to provide remedial instruction when they are already dedicating a fair amount of time to meeting their state standards in all content areas. Furthermore, parents and students do not want to be deprived of all opportunities for school engagement in activities such as art, music, athletics, and band.

Purpose and rationale

The purpose of this study was to investigate the relative effects of a yearlong intervention varying group size with older students (seventh- and eighth-graders) with reading difficulties. Students with reading difficulties were randomly assigned to one of three conditions: (a) research small-group treatment with teacher hired and trained by research team (e.g., one teacher and approximately five students), (b) research large-group treatment with teacher hired and trained by research team (e.g., one teacher and ~12–15 students), and (c) school comparison group (school-delivered instruction in groups of ~12–15 students). Students in the treatment conditions (i.e., research large group and research small group) were provided the same multi-component instructional intervention that addressed multi-syllable word reading, academic vocabulary acquisition, reading fluency, and comprehension. This study was conducted daily for 50 min over the course of a full school year as part of a large-scale, multiyear investigation with older students (Vaughn et al., 2008).

Middle schools provide unique challenges to implementing effective interventions for students with reading difficulties. Unlike elementary school educators, who have a long history of providing reading interventions to students with reading difficulties, middle school educators have typically expected students to be able to read well enough to learn from text. Recent studies report that this expectation is no longer realistic and that the number of students at the middle school level requiring additional instruction in reading ranges from 20 to 40%, depending upon the district (Lee, Grigg, & Donahue, 2007). A fundamental question underlying effective interventions for secondary students is how to accelerate student progress within the framework of secondary school schedules and with limited resources. Providing sufficiently intensive interventions requires well-trained teachers, along with extended time for instruction and small class sizes. In this study, we systematically investigated the effects of varying group size on the acceleration of students' reading progress. While it is reasonable to think that very small instructional groups would be most likely to make an impact on reading outcomes, we are also cognizant of the realities of implementing interventions for older students within the scheduling and course requirements of secondary settings. Thus, this study was designed to accomplish two important aims: (a) determine the effectiveness of a multi-component reading intervention by comparing outcomes for treatment and comparison students and (b) determine the relative effects of group size by systematically varying group size and holding intervention constant.

Method

Participants

This study reports on 546 seventh- and eighth-grade struggling readers during the 2006–2007 academic year. Descriptions of the criteria for student participation and sample selection, including the school sites, are described below.

School sites

This study was conducted in the southwestern United States in two large, urban cities. Approximately half of the sample was recruited from each site. Seventh- and eighth-grade students from six middle schools (two from the first and larger site and four from the second site) participated in the study. From the first site, the two schools were classified as urban; the remaining four schools were classified as suburban and rural, all with school populations ranging in size from 633 to 1,300 students. The rate of students qualifying for reduced 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 sampled from students in grades 7–8. Struggling readers were identified based on their performance on the state accountability test, the Texas assessment of knowledge and skills (TAKS), an assessment of reading comprehension. We identified struggling readers as students who did not pass TAKS (performance below 2,100 standard score) or whose test score was within one-half of one standard error of measurement above the passing criteria (performance within 2,100–2,150 standard scores) on the first administration of TAKS in the spring of the previous school year. These "bubble" students have scores close enough to the passing standard that they could fail the TAKS test in the future because of measurement error. We also identified students exempted from TAKS who took the school determined alternative assessment (SDAA), a test designed for special education students with very low reading performance.

The struggling readers were randomly assigned within school to either treatment or business as usual intervention in a 2:1 ratio. Students were excluded from participation if: (a) they were enrolled in a life skills class; (b) their SDAA performance levels were below 3.0; (c) they had a significant sensory or developmental disability (e.g., blindness, deafness, autism) or serious behavioral/ emotional disorder; or (d) they were enrolled in an English as a second language class.

Student participants

The struggling readers were selected from a larger sample of 3,815 grade 7–8 grade students in the spring of the 2005–2006 academic year. Students who met criteria were
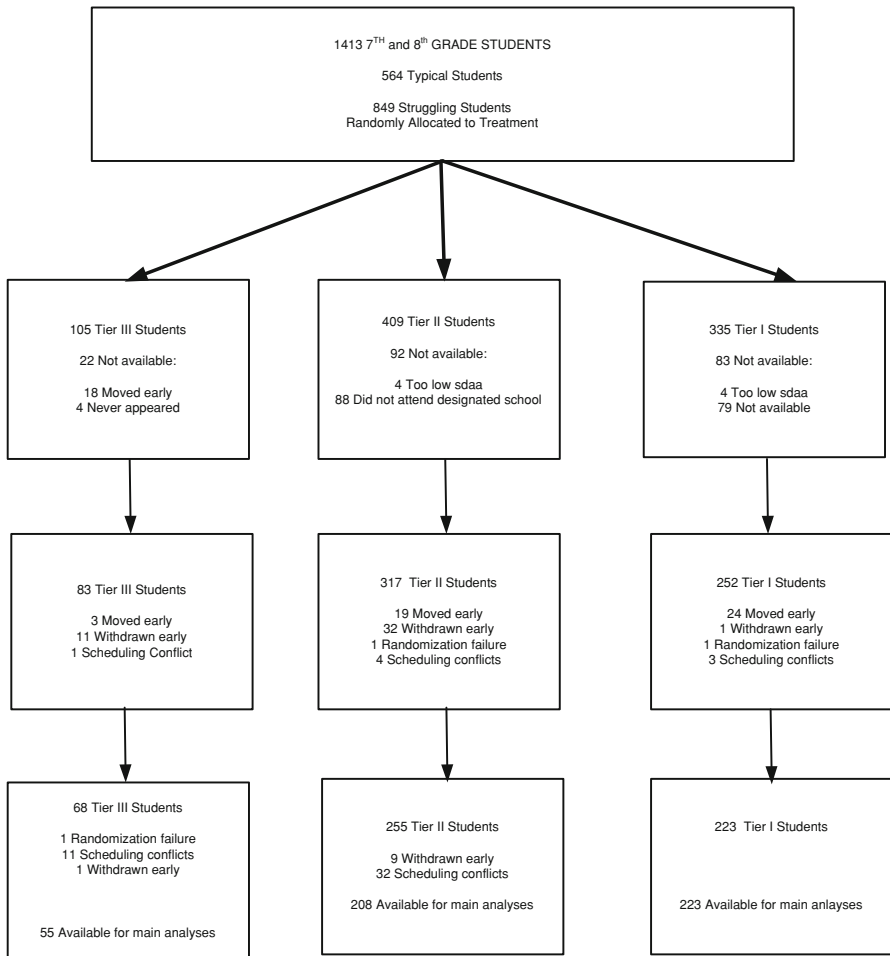
Fig. 1 Randomization by group

randomized to one of three conditions, as outlined in Fig. 1 (i.e., research small-group treatment, research large-group treatment, or school treatment comparison). Each condition was allotted 25 instructional groups. Group size was defined as 3–5 students for the research small-group treatment, 10–15 students for the research large-group treatment, and 10–20 students for the school comparison condition (i.e., English language arts or reading). Because the treatment is group size, a greater number of students were randomized to the research large-group treatment and business as usual conditions than the research small-group treatment (see Fig. 1).

Figure 1 shows that 409 students were randomized to receive research large-group treatment; 105 students were randomized to receive research small-group treatment; and 335 students were randomized to the school comparison condition. Because students' status as a struggling reader was based on Spring 2006 TAKS (prior to treatment), randomization occurred at that time. Some of these students did

not attend the middle school their grade 6–7 feeder pattern data suggested, and the proportion did not differ across the randomized struggling reader groups. Specifically, this situation evidence for 92 (23%) randomized to the research large-group treatment, 22 (21%) randomized to the small-group treatment, and 83 (25%) randomized to the school comparison condition.

The remaining sample of 652 students included 83 students in research small group, with 55 available at posttest; 317 students in research large group, with 208 available at posttest; and 252 students in the school comparison condition, with 223 available at posttest. Fifteen students in the research small-group treatment, 55 students in the research large-group treatment, and 29 students in the school comparison condition were tested at pretest but were not followed further because of early withdrawals or moves from school, conflicts that prevented participation in the study, or randomization issues.

Following pretest, a further 13 students in the research small-group treatment and 47 students in the research large-group treatment were not present in their schools at the end of the school year. From randomization at the beginning of the year to posttest at the end of the year, a total of 48% of the students in the research small-group treatment, 48% of the students in the research large-group treatment, and 33% of the students in the school comparison condition left the study. For students who left the study, performance on pretest measures did not differ significantly from that of students who remained in the study ($p > 0.05$).

Intent-to-treat analyses were performed on the 255 students in the research large-group treatment, 68 students in the research small-group treatment, and 223 students in the school comparison condition available at posttest. On-treatment analyses of 208 large-group treatment students, 55 research small-group treatment students, and 223 school comparison students were also conducted. Given that the intent-to-treat analyses closely paralleled the on-treatment analyses and that the students who did not receive intervention did not differ at pretest from those who did receive intervention, the remainder of this report focuses only on the 208 research large-group treatment and 55 research small-group treatment students who actually received the intervention and the 223 school comparison students.

Each school contributed between 22 and 186 students to the sample of 486 students. Forty-three percent of the sample was female, and 74% of the sample qualified for free or reduced lunch (2.3% did not provide data on free or reduced lunch). One hundred ninety-three students (40%) were African American, 211 (43%) were Hispanic, 67 (14%) were Caucasian, 13 (3%) were Asian, and 2 (0.41%) were American Indian. The proportions of students from the treatments did not differ in terms of site, sex, free or reduced lunch status, or ethnicity (all $p > 0.05$). Students in the two struggling reader groups also did not differ in terms of their struggling reader category (e.g., failure, bubble, or special education) or in age ($p > 0.05$).

Intervention instructors for both research treatments: small and large group

Fifteen certified teachers hired and trained by the research team provided intervention for the research small-group and research large-group treatments. Teachers had an average of 6.3 years of teaching experience. All teachers had an

undergraduate degree, and ten teachers had a master's degree in education. Twelve of the 15 teachers had a teaching certification in a reading-related field.

The investigators provided intervention teachers with ∼60 h of professional development prior to teaching as well as an additional 9 h of professional development related to the intervention throughout the remainder of the year. Professional development sessions included training on intervention specific methods, features of effective instruction, behavior management, and general information about the adolescent struggling reader. Teachers also received ongoing feedback and coaching and participated in bi-weekly staff development meetings.

Description of intervention

All classroom teachers participated in a Tier 1 intervention addressing comprehension and vocabulary instruction in content area classrooms. Details concerning the professional development program can be found elsewhere (Denton, Bryan, Wexler, Reed, & Vaughn, 2007). For the interventions, the instruction took place for 45–50 min per day (regular class period) throughout the school year (September–May).

Three phases of instruction, varying in emphasis, composed the yearlong intervention. Each phase prioritized a selected element of instruction (e.g., comprehension through extended text reading), and the skills and strategies taught in previous phases were supported in the subsequent phases. In Phase I (∼7–8 weeks), word study and fluency were emphasized, with additional instruction provided in vocabulary and comprehension. Phase II (∼17–18 weeks) emphasized vocabulary and comprehension, with additional instruction and practice in the application of the word study and fluency skills and strategies learned in Phase I. Phase III (8–10 weeks) continued the instructional emphasis on vocabulary and comprehension, with more time spent on independent student application of skills and strategies. The components of instruction included in each phase are described in detail below.

Phase I

Fluency

Higher- and lower-performing readers were paired for reading fluency practice. Student pairs engaged in daily repeated reading with error correction and graphing of the number of words read per minute. Each student read passages three times, with the more proficient reader in the pair always reading first. The teacher circulated and provided feedback to students on the partner reading procedures, modeled fluent reading, and provided feedback on students' fluency.

Word study

Teachers used the lessons in REWARDS Intermediate (Archer, Gleason, & Vachon, 2005) to teach advanced word study strategies for decoding multisyllabic words. Daily instruction and practice was provided in individual letter sounds, letter

combinations, affixes, applying a strategy for decoding multisyllabic words by breaking the words into known word parts, spelling multisyllabic words, and applying these skills and strategies while reading expository text.

Vocabulary

Unfamiliar vocabulary words were directly taught, and teachers and students maintained vocabulary word walls and charts to review and use previously taught vocabulary. Words were selected from text reading and/or words that were used to teach the multisyllabic word reading application described above. Simple definitions of the words along with examples/nonexamples of how to use the words were provided. Students then practiced applying knowledge of the new words by identifying appropriate use of the words and matching the new words to examples of how the words could be used.

Comprehension

Students engaged in daily discussion of the passages read in the word study component. Literal and inferential questions regarding the passage were discussed during and after the reading. Students were taught to locate information in text, including rereading when necessary, to develop answers to the questions.

Phase II

Word study and vocabulary

The skills and strategies learned in Phase I were reviewed daily in Phase II through application to reading and spelling new vocabulary words and reading connected text. After reading new vocabulary words, students were provided with simple definitions for each word and practiced using their knowledge of the word by matching appropriate words to various scenarios, e.g., or descriptions. In addition, students were introduced to word relatives and their parts of speech (e.g., *politics*, *politician*, *politically*). Vocabulary words for instruction were chosen from the text read in the fluency and comprehension component of the lesson.

Fluency and comprehension

The majority of instructional time in Phase II was spent in reading and comprehending connected text. Teachers utilized two sources for text and fluency/comprehension lessons: (a) expository text from REWARDS Plus Social Studies (Archer et al., 2005) and (b) narrative text from reading-level appropriate novels. In each lesson, teachers introduced background knowledge necessary to understand the text. Students then read the text at least twice for fluency. During and after the second reading, students engaged in discussion of the text through questioning. In addition, teachers provided explicit instruction in generating questions while reading. Students learned to generate literal questions, questions

requiring the synthesis of information in text, and questions requiring the application of background knowledge to information in text. Instruction in strategies for identifying the main idea; summarizing; and addressing multiple-choice, short-answer, and essay questions was also provided in Phase II.

Phase III

Word study and vocabulary

The word study and vocabulary instruction in Phase III was identical to Phase II. However, teachers used fluency and word reading activities and novel units developed by the research team. Word study exercises were based on words selected from these materials.

Fluency and comprehension

Students were provided daily review and practice of previously learned skills and strategies. As in Phase II, both expository and narrative (novels) texts were used for instruction. The focus of Phase III was on application of the previously learned strategies to independent reading. Students read passages, generated questions about the text, and addressed comprehension questions related to all the skills and strategies learned (multiple choice, main idea, summarizing, literal information, synthesizing questions, background knowledge, etc.) independently before discussing. Teachers provided corrective feedback and reteaching as necessary.

Fidelity of implementation

Implementation validity checklists were completed for each teacher monthly. A three-point Likert-type scale ranging from 1 (low) to 3 (high) was used to rate the implementation of each treatment component as well as the quality of implementation. Implementation ratings assessed the extent to which teachers completed the required elements and procedures to meet the objectives of each component. Quality of implementation ratings assessed the active engagement of students during the learning activities with frequent student responses and the appropriateness of instruction, including monitoring student performance, providing feedback, and adjusting pacing. A total score for fidelity of implementation was calculated by averaging the implementation and quality ratings across all of the instructional components and observations.

Two observers collected the fidelity of implementation data across the year. The observers rotated the teachers observed each month so that each teacher was seen by each observer several times. Prior to data collection, the two observers were trained on the observation measure. After training, the two observers completed the fidelity measure independently for two different teachers and classes. Interrater reliability was 100% on the first observation and 93% on the second observation. At midyear, the two observers again completed the fidelity measure independently for one class, with a reliability of 94%.

Average implementation scores for teachers ranged from 2.22 to 3.00. Average quality of instruction scores ranged from 2.00 to 3.00. Total fidelity of intervention implementation (all implementation and quality ratings averaged across observations) ranged from 2.45 to 2.96.

Additional intervention

We were also interested in the extent to which students received additional reading intervention from their schools. Student schedules were obtained in December and May to determine any additional reading intervention the struggling readers in the small-group and large-group treatments and the business as usual groups were receiving. When a student was receiving additional reading intervention, the teacher was interviewed to obtain information about the intervention.

Of the 208 students in the large-group treatment, 170 (82%) reported receiving no additional instruction. Of the remainder, 38 (17%) reported receiving one additional type of instruction, and four students (0.02%) received two. The type of supplemental instruction varied, but included tutoring (without a specific program name), Fundamentals of Reading, resource classes, reading enrichment, SPARK, Read 180, and state accountability test tutoring. Certified teachers nearly always delivered additional instruction. Group size varied from 2 to 16 students but was most often provided in groups of 10 to 15 students. The average amount of time of additional instruction for the 38 large-group treatment students who received additional intervention was 127.2 h (SD = 53.5, range 24.0–255.0, median = 141.7). The four students receiving a second form of additional instruction received, on average, an additional 72.9 h of instruction (SD = 2.4, range 70.8–75.0, median = 72.9).

Of the 55 research small-group treatment students, 42 (76%) reported receiving no additional instruction. Of the remainder, 13 (33%) had reports of receiving one additional type of instruction, and one student received two. The type of supplemental instruction varied, but largely consisted of resource classes or SPARK. A certified teacher nearly always delivered this supplemental instruction, typically in groups of 6–15 students. The average amount of time of additional instruction provided to 13 research small-group students was 97.6 h (SD = 34.8, range 50.0–141.7, median = 111.0). The student receiving a second form of additional instruction received 67.5 h.

Of the 223 school comparison students, 168 (75%) reported receiving no additional instruction. Of the remainder, 55 (25%) reported receiving one additional type of instruction and four students reported receiving two. The type of supplemental instruction varied, but largely consisted of Reading Enrichment, state accountability test tutoring, unspecified tutoring, Read 180, English as a second language, and resource classes. A certified teacher nearly always delivered this supplemental instruction, typically in groups of 10–15 students. The average amount of time of additional instruction for these students was 124.5 h (SD = 62.3, range 7.0–277.5, median = 127.5). For the students receiving a second type of additional instruction, the average amount of time was 103.1 h (SD = 56.3, range = 75.0–187.5, median = 75.0). A greater proportion of students in the school comparison condition received additional instruction (47%) relative to those in the

large-group treatment (25%) and those in the small-group treatment (33%; $p < 0.05$ overall). The proportion of students receiving additional instruction was similar across sites ($p > 0.05$). In terms of *amount* of this additional instruction, within the total group, as well as in the smaller group of students who received additional instruction, there was not an interaction of site and treatment group (all $p > 0.05$). A formal evaluation revealed that additional instruction did not substantively change the interpretation of any treatment effects that were present.

Measures

Texas assessment of knowledge and skills

The TAKS (Texas Education Agency [TEA], 2004a, b, c) is a criterion-referenced assessment specific for each grade that is aligned with state grade-based standards. Students read both expository and narrative passages and then answer several multiple-choice/short-answer questions designed to assess the literal meaning of the passage, vocabulary, and different aspects of critical reasoning about the material read. The internal consistency (coefficient alpha) of the grade 7 test is 0.89 (TEA, 2004a, b, c). A variety of studies have found excellent construct validity comparing student performance on TAKS with other assessments, such as the national assessment of educational progress (NAEP) and the norm-referenced Iowa Tests, college readiness measures (TEA, 2004a, b, c), as well as individual norm-referenced assessments (Denton, Wexler, Vaughn, & Bryan, 2008).

Group reading assessment and diagnostic evaluation

At both pretest and posttest, students were administered the Passage Comprehension and Listening Comprehension subtests of the group reading assessment and diagnostic evaluation (GRADE; Wilder & Williams, 2001; Form A). The GRADE is a group-based, norm-referenced, untimed test. For Passage Comprehension, the students read five to six narrative or expository excerpts and answer multiple-choice questions that require questioning, predicting, clarifying, and summarizing text. For Listening Comprehension, the examiner reads aloud one or two sentences, and the student marks one of four pictures that conveys the meaning of what the examiner read. The GRADE produces a stanine score for the Passage Comprehension subtest, but for purposes of this study, the raw score was prorated to derive a standard score for the GRADE Comprehension Composite, which is typically based on the Passage Comprehension and Sentence Comprehension measures (the latter was not administered). Coefficient alpha for the Passage Comprehension subtest in the entire sample of 486 struggling readers and 440 typical readers who contributed data throughout the year was 0.87 at pretest time point.

Woodcock-Johnson III tests of achievement

At both pretest and posttest, students were individually administered the Letter-Word Identification, Word Attack, and Passage Comprehension subtests of the

Woodcock-Johnson III Tests of Achievement (WJ-III; Woodcock, McGrew, & Mather, 2001), except for Spelling, which was administered in a group posttest. The Letter-Word Identification subtest assesses the ability to read real words, with a median reliability of 0.91. The Word Attack subtest examines the ability to apply phonic and structural analysis skills to the reading of nonwords, with a median reliability of 0.87. The Passage Comprehension subtest utilizes a cloze procedure to assess sentence-level comprehension by requiring the student to read a sentence or short passage and fill in missing words based on the overall context. The Passage Comprehension subtest has a median reliability of 0.83. The Spelling subtest involves orally dictated words written by the examinee to assess encoding skills, which are related to decoding ability; this measure was modified for group administration by administering a set list of items. The Spelling subtest has a median reliability of 0.83. Standard scores from these subtests were the dependent measures of interest. Coefficient alphas in the entire sample of 486 struggling readers and 440 typical readers who contributed data throughout the year for the Letter-Word Identification, Word Attack, and Passage Comprehension subtests at pretest were 0.98, 0.94, and 0.96, respectively, and at posttest were 0.97, 0.99, and 0.93, respectively; coefficient alpha for Spelling at posttest was 0.94.

Test of sentence reading efficiency

At both pretest and posttest plus three additional time points, students were administered the test of sentence reading efficiency (TOSRE; Wagner, in press). The TOSRE is a 3 min, group-based assessment of reading fluency and comprehension. Students are presented with a series of short sentences and are required to determine whether they are true or false. The TOSRE was standardized on 2,000 students from grades 4–9. The standard score was the dependent measure utilized. The mean intercorrelation of performances across the five time points in the entire sample of 486 struggling readers and 440 typical readers was 0.96 for standard scores and 0.96 for raw scores. These correlations likely underestimate reliability because some students received intervention and may have changed their rank order over time.

AIMSweb reading maze

At pretest and posttest plus three additional time points during the year, the Maze subtest of the AIMSweb (Shinn & Shinn, 2002) was administered. The AIMSweb Maze is a 3 min, group-administered assessment of fluency and comprehension. Each AIMSweb Maze is a 150- to 400-word passage, in which the first sentence is intact but every seventh word thereafter is deleted. Students are required to identify a correct target word from among three choices for each missing word. The raw score is the number of targets correctly identified within the 3 min time limit and was the dependent measure utilized. AIMSweb provides 15 different stories for seventh and eighth grade, and the particular story any individual student received was randomly determined within grade, school, and treatment group. These measures are not equated, although stories were chosen based on reading level. The

mean intercorrelation of performances across the five time points in the entire sample of 486 struggling readers and 440 typical readers was 0.95.

Test of word reading efficiency

At both pretest and posttest, the Sight Word Efficiency and Phonemic Decoding Efficiency subtests of the test of word reading efficiency (TOWRE; Torgesen et al., 1999) were administered. For the Sight Word Efficiency subtest, the participant is given a list of 104 words and asked to read them as accurately and as quickly as possible; the number of words read correctly within 45 s is recorded. For the Phonemic Decoding Efficiency subtest, the participant is given a list of 63 nonwords and is asked to read them as accurately and as quickly as possible; the number of nonwords read correctly within 45 s is recorded. Psychometric properties are good, with most alternate forms and test retest reliability coefficients at or above 0.90 in this age range (Torgesen et al., 1999). Standard scores from these subtests were the dependent measures.

Passage fluency

At pretest and posttest plus three additional time points, the passage fluency (PF) was administered. The PF was designed by the authors to measure oral reading fluency. The PF consists of 100 graded passages (50 narrative and 50 expository) for use in grades 6–8 that are administered as short, 1 min probes. All passages averaged ~ 500 words each and ranged in difficulty from 350 to 1,400 Lexiles (Lexile Framework, 2007). The 100 passages were placed into ten "Lexile bands" separated by 110 Lexile units. Each Lexile band comprised ten passages, five of which expository and five narrative.

   Students were administered the PF at five time points throughout the year. At each time point, students read five stories for 1 min each. Each of the passages was selected from a different Lexile band, with two selected below grade level, one at grade level, and two above grade level. The overall difficulty range for each student was 550 Lexiles. The particular stories any individual student received were randomly determined within school, grade, and treatment group. For purposes of this report, the dependent measure utilized is a linearly equated-score average of the five 1 min probes based on a larger sample of 1,803 middle school students in grades 6–8. Equating was carried out within grade and time point. The equated scores eliminated differences between stories in mean differences and in within-story variability at each assessment time point, but allowed differences over time and across grade in both mean performance and variability in performance. Therefore, differences in mean performance across time points and grades are preserved, but any resulting differences are not due to, for e.g., older students reading easier passages or students reading difficult passages followed by reading easier passages later in the year.

Word list fluency

The word list fluency (WLF) was also designed by our research team to assess decontextualized word reading fluency on lists that vary in difficulty. Students are required to read as many words as possible within 1 min for three word lists (two constructed lists and one passage list). The first pool, constructed lists, comprised 21 timed lists subdivided into seven "easy" lists, seven lists of "moderate" difficulty, and seven "challenging" word lists. These words were derived from a word frequency guide (Zeno, Ivens, Millard, & Duvvuri, 1995) word bank, and each list comprised ∼150 words, constructed on the basis of word length and frequency parameters. For the purposes of this study, "easy" words were less than five letters in length and were considered high-frequency words. The individual word lists were constructed from a random sample of short, high-frequency words. Words for the "moderate" word lists were constructed in a similar manner, with the word length parameter set at 6–10 letters and the frequency parameter remaining high. The word length parameter for "challenging" word lists remained at 6–10 letters, but the frequency parameter was adjusted to low. The second pool, passage lists, consisted of 38 word lists derived from the PF measure described above. For each story, unique words were identified and the duplicates removed. The unique words were then randomly ordered (within passage) and arranged into a word list.

At each administration time point, students read three word lists of varying difficulty for 1 min each. Students were randomly assigned to read one of three types of word list described above: (a) passage word lists derived from the *same* stories they read to assess PF, (b) passage word lists derived from stories that *other* students read to assess PF (but they themselves did not read), or (c) constructed word lists. Students reading constructed lists read one easy, one moderate, and one challenging list. Students reading passage lists read two lists comprising words from stories considered below grade level, one list comprising words from stories considered at grade level, and two lists comprising words from stories considered above grade level. Again, the particular stories and, thus, the particular lists that any individual student received were randomly determined within school, grade, and treatment group. As with PF, the dependent measure utilized for WLF is 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 486 struggling readers and 440 typical readers was 0.97 at pretest and 0.98 at posttest.

Kaufman brief intelligence test—2

For descriptive purposes, both the Matrices and Verbal Knowledge subtests of the Kaufman brief intelligence test—2 (K-BIT 2; Kaufman & Kaufman, 2004) were administered. Internal consistency values for the subtests and composite range from 0.87 to 0.95, and test–retest reliabilities range from 0.80 to 0.95, in the age range of the students in this study (Kaufman & Kaufman, 2004). The Matrices subtest, a measure of nonverbal problem solving, was administered at pretest. The Verbal Knowledge subtest, which assesses receptive vocabulary and general information (e.g., nature, geography), was administered at posttest. The K-BIT 2 Riddles subtest

was not utilized; therefore, the Verbal Knowledge score was prorated for the verbal domain.

Analysis plan and preliminary decisions

Analyses were conducted in the context of the generalized linear model (Dobson, 1990; Green & Silverman, 1994) using SAS (SAS Institute, 2002–2003) and MPLUS v. 5 (Muthen & Muthen, 1998–2007). For measures with two time points, analysis of covariance (ANCOVA) models were fit with pretest performance as the primary covariate. School site, the presence/absence of additional instruction, fidelity, intervention time, students' age, and intervention group size were moderators of particular interest. Specifically, we were interested in whether (high) fidelity, (more) intervention time, and (smaller) group size were related to outcomes of interest.

The nested structure of the data was considered when the primary models indicated a significant treatment main effect or when the raw effect size was greater than $+0.15$ favoring the treatment group, except where otherwise noted. We addressed nesting factors in several ways: by classroom reading teacher, by tutoring group for treatment students, or some combination of these, evaluating the extent to which clustering at the teacher level explained significant variability in outcomes. In general, the effect of clustering was low (below 10%) and not significant ($z > 0.05$), regardless of its structure. The exception was the effect of intervention group on measures of spelling and comprehension, particularly TAKS and GRADE, and these models were adjusted for the influence of within-group correlation.

Results

With several exceptions, data were normally distributed. Six of 11 variables exhibited skewness or kurtosis greater than |1| at pretest, with outliers greater than three standard deviations. A similar, though more subtle, pattern was noted for posttest distributions.

Outcomes

Pretest and posttest Ns, means, standard deviations, statistical tests, and unadjusted effect sizes are presented in Table 1 for the on-treatment group of struggling readers with data available at both the beginning and ending of the year. Performance is presented by group, and variables are organized into measures of decoding and spelling, comprehension, and fluency. Because students were successfully randomly assigned to condition, there were no pretreatment differences across the three groups. Age was related to all outcomes (all ps < 0.05, range 0.003–0.903, median = 0.066); its inclusion as a covariate provided additional power for evaluating treatment effects. Pretest results were similar when comparisons between treatment groups were made with all students available at pretest (as opposed to those with both pretest and posttest data, as in Table 1).

Table 1 Pretest and posttest performance on reading measures for treatment groups

Measure	N	Group	Pre M	Pre SD	Post M	Post SD	Comparison	d	CI	F	p
TAKS	171	BAS	2,034.2	94.7	2,134.2	94.7					
	153	Large	2,028.4	100.7	2,105.9	151.7	BAS vs. Large	0.19	−0.27 to +0.410	1.54	0.28
	44	Small	2,044.7	86.6	2,120.7	143.5	BAS vs. Small	0.09	−0.237 to +0.426	1.54	0.97
GRADE reading comp	218	BAS	87.2	11.4	86.5	10.0					
	203	Large	86.3	10.5	86.1	10.9	BAS vs. Large	0.04	−0.153 to +0.229	0.07	1.0
	55	Small	85.6	9.3	86.2	8.4	BAS vs. Small	0.03	−0.263 to +0.328	0.07	1.0
WJ-III passage comp	209	BAS	85.5	11.3	86.8	12.2					
	194	Large	85.1	11.9	86.7	11.7	BAS vs. Large	0.01	−0.187 to +0.203	0.75	1.0
	50	Small	84.9	9.4	87.7	9.4	BAS vs. Small	0.08	−0.226 to +0.391	0.75	0.67
AIMSweb maze	218	BAS	16.7	7.9	25.1	9.8					
	202	Large	16.8	7.6	24.5	9.9	BAS vs. Large	0.06	−0.131 to +0.252	0.38	1.0
	55	Small	15.5	6.6	23.5	9.3	BAS vs. Small	0.17	−0.129 to +0.463	0.38	1.0
TOSRE	218	BAS	80.6	14.2	86.8	15.9					
	203	Large	80.6	13.0	85.1	15.4	BAS vs. Large	0.11	−0.83 to +0.300	1.41	0.37
	55	Small	79.5	13.3	86.3	15.5	BAS vs. Small	0.03	−0.264 to +0.328	1.41	1.0
Passage fluency	213	BAS	114.1	36.3	137.8	37.9					
	194	Large	115.1	30.9	141.5	36.2	BAS vs. Large	0.10	−0.209 to +0.408	2.7	0.35
	50	Small	121.9	31.1	151.0	34.3	BAS vs. Small	0.36	−0.055 to +0.674	2.7	0.11
List fluency	213	BAS	76.1	27.9	82.0	28.6					
	194	Large	74.1	25.8	83.8	27.1	BAS vs. Large	0.07	−0.130 to +0.259	4.1	0.01
	50	Small	78.2	21.9	86.2	26.4	BAS vs. Small	0.15	−0.156 to +0.461	4.1	0.80
TOWRE SWE	209	BAS	90.8	11.9	91.5	12.8					
	194	Large	91.0	10.4	92.4	11.9	BAS vs. Large	0.07	−0.123 to +0.268	0.15	0.98
	50	Small	92.8	10.4	95.1	12.3	BAS vs. Small	0.29	−0.026 to +0.596	0.15	0.32

Table 1 continued

Measure	N	Group	Pre M	Pre SD	Post M	Post SD	Comparison	d	CI	F	p
TOWRE PDE	209	BAS	93.6	16.3	92.8	15.3					
	194	Large	93.1	14.1	93.7	14.6	BAS vs. Large	0.06	−0.135 to +0.256	1.17	0.39
	50	Small	96.8	14.5	95.7	14.9	BAS vs. Small	0.19	−0.118 to +0.501	1.17	1.0
WJ-III letter-word ID	209	BAS	91.6	13.5	92.1	12.5					
	194	Large	90.9	10.2	91.4	12.2	BAS vs. Large	0.06	−0.139 to +0.252	0.17	1.0
	50	Small	93.3	10.2	94.0	9.3	BAS vs. Small	0.17	−0.137 to +0.481	0.17	1.0
WJ-III word attack	209	BAS	95.2	11.7	94.8	11.4					
	194	Large	94.6	11.1	95.2	11.5	BAS vs. Large	0.04	−0.161 to +0.230	0.77	0.81
	50	Small	95.8	8.6	95.8	8.8	BAS vs. Small	0.10	−0.211 to +0.407	0.77	1.0
WJ-III spelling[a]	208	BAS	91.7	13.5	91.8	14.4					
	193	Large	90.8	11.8	91.9	13.7	BAS vs. Large	0.01	−0.304 to +0.318	0.66	1.0
	49	Small	93.5	10.2	95.0	11.9	BAS vs. Small	0.24	−0.072 to +0.552	0.66	0.81

BAS business as usual group; *Large* large-group intervention; *Small* small-group intervention; *Pre M* mean of pretest performance; *Pre SD* standard deviation of pretest performance; *CI* confidence interval of effect size; *F* statistical test of the difference between adjusted means of BAS and Large subgroups or BAS and Small subgroups ($df[1, N-3]$ where *N* is BAS + Large or BAS + Small); *TAKS* Texas Assessment of Knowledge and Skills; *GRADE Reading Comp* GRADE Reading Comprehension subtest; *WJ-III Passage Comp* Woodcock Johnson III Passage Comprehension; *AIMSweb Maze* AIMSweb Maze Reading Comprehension; *TOSRE* Test of Sentence Reading Efficiency; *Passage Fluency* investigator-created measure of oral reading fluency of connected text; *List Fluency* investigator-created measure of word reading fluency; *TOWRE SWE* Test of Word Reading Efficiency Sight Word Efficiency subtest; *TOWRE PDE* Test of Word Reading Efficiency Phonemic Decoding Efficiency subtest; *WJ-III Letter-Word ID* Woodcock-Johnson III Letter-Word Identification subtest; *WJ-III Word Attack* Woodcock Johnson-III Word Attack subtest; *WJ-III Spelling* Woodcock Johnson-III Spelling subtest. *F* values shown are the centered treatment effect in the context of the observed interaction of pretest and treatment group

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

Decoding and spelling

The WJ-III Letter-Word Identification, Word Attack, and Spelling subtests were the primary measures of decoding and spelling. For the WJ-III Letter-Word Identification subtest, there was a three-way interaction of pretest, treatment, and site, $F(2, 464) = 4.03, p = 0.02\ 0.46, \eta_p^2 = 0.18$. Further analyses examining the relation of pretest performance and treatment condition within each site indicated that the interaction and main treatment effect was not significant. However, the pattern of effects was in the expected direction, with an effect size of 0.06 favoring large-group treatment (difference between large-group treatment and business as usual) and 0.17 favoring small-group treatment (difference between small-group intervention and business as usual). Also, within the smaller site, there was a three-way interaction among pretest, age, and treatment condition, $F(2, 262) = 2.80, p = 0.04, \eta_p^2 = 0.03$. To further analyze the effect of age, the three treatment groups were plotted separately, with age dichotomized into two groups (i.e., older and younger students). The pattern of findings indicated that the prediction was flatter for younger than older students but that age had little impact on prediction. For the WJ-III Word Attack subtest, the main effect of treatment was not significant, $F(2, 449) < 1, p < 0.46$, $\eta_p^2 = 0.003$, with effects of 0.04 for large-group treatment and 0.10 for small-group treatment. WJ-III Spelling was administered at posttest only, and WJ-III Letter-Word Identification was used as the pretest covariate. The main treatment effect was not significant, $F(2, 446) = 69.5, p < 0.52, \eta_p^2 = 0.003$, and the effect sizes were 0.01 and 0.24 for large-group and small-group treatments, respectively. There were no significant moderating effects such as presence/absence of additional instruction, fidelity, intervention time, age, or intervention group size for decoding and spelling outcomes. Cluster-related effects were not evaluated (Table 2).

Comprehension

The TAKS, GRADE, and WJ-III Passage Comprehension subtest were the primary measures of reading comprehension. School officials administered the TAKS ~3 months earlier than other measures used in this study. Also, students who qualified for SDAA in either year were not represented in these analyses. As a result, sample sizes reported for the TAKS measure tend to be smaller than samples for the other outcomes. There were no significant treatment effects for TAKS, $F(1, 364) = 1.54$, and trends in the group-specific effect sizes were contrary to expectations (0.19 for large group and 0.09 for small group), though this may be an artifact of the differing levels of the test, error in the estimates of effect, or a combination of these. We also evaluated the extent to which treatment differentially increased the chances of passing the TAKS test from year to year. Among those who took TAKS in 2007, students who received large-group instruction were more likely to pass (76%) relative to those in the business as usual group (57%), $\chi^2(df = 1, N = 96) = 4.03, p < 0.05$.

For the GRADE, there was a three-way interaction of pretest, site, and treatment condition, $F(2, 464) = 4.03, p = 0.018, \eta_p^2 = 0.017$. The relation among pretest and treatment condition was further examined within site. Within the larger site, the

Table 2 Fidelity of implementation

Teacher	Average implementation rating	Average quality rating	Total fidelity of implementation
1	2.93	2.71	2.81
2[a]	2.75	2.25	2.50
3	2.33	2.82	2.55
4	2.59	2.69	2.64
5	2.22	2.73	2.45
6	3.00	2.93	2.96
7[b]	3.00	2.00	2.50
8[c]	2.33	2.60	2.45
9	2.76	2.88	2.82
10	3.00	2.67	2.83
11[b]	2.69	2.62	2.65
12[c]	2.57	3.00	2.79
13[a]	2.73	2.82	2.77

[a] Teachers 2 and 13 taught the same classes. Teacher 2 taught the class for ∼2 months

[b] Teachers 7 and 11 taught the same classes. Teacher 7 taught in the fall. Teacher 11 continued the classes in the spring

[c] Teachers 8 and 12 taught the same classes. Teacher 12 taught the first half of the year and Teacher 8 continued the classes in the spring

interaction of pretest and treatment condition and main effect of treatment condition was not significant. Within the smaller site, slopes for the large-group intervention were steeper than business as usual, $p = 0.039$, and slopes of large-group intervention were steeper than small-group intervention, $p = 0.02$. The estimates of effect were 0.04 for large group and 0.03 for small group. Results on the WJ-III Passage Comprehension subtest revealed no main effects, $F(2, 449) = 0.75$, $p < 0.47$, and negligible effect estimates (0.01 and 0.08 for large-group and small-group treatments, respectively). There were no effects related to moderation (presence/absence of additional instruction, fidelity, intervention, time, students' age, and intervention group size) or clustering.

Fluency

The TOWRE Sight Word Efficiency subtest, TOWRE Phonemic Decoding Efficiency subtest, AIMSweb Mazes, TOSRE, Passage Fluency, and Word List Fluency were the primary measures of reading fluency. For the TOWRE Sight Word Efficiency, there was a three-way interaction of pretest, treatment condition, and age, $F(3, 432) = 5.13$, $p = 0.0002$, $\eta_p^2 = 0.034$; the effect sizes were 0.07 for large group and 0.29 for small group. To further examine the effect of age, the three treatment conditions were plotted separately, with age dichotomized into two groups (i.e., older and younger). The pattern of findings indicated that the prediction was flatter for younger than older but that age had little impact on prediction in terms of variance accounted for. The pattern of effects was in the expected direction, with an effect size of 0.11 favoring large-group treatment (difference between large-group treatment and business as usual) and 0.03 favoring small-group treatment (difference between small-group intervention and business as usual). No significant moderator or clustering effects were identified.

For TOWRE Phonemic Decoding Efficiency, there was no overall difference in adjusted means, $F(2, 449) = 1.17$, $p > 0.05$, $\eta_p^2 = 0.005$. The pattern of effects was in the expected direction, with an effect size of 0.06 favoring large-group treatment and 0.19 favoring small-group treatment. No significant moderator or clustering effects were identified.

For AIMSweb Mazes, there was a three-way interaction of pretest, treatment group and site, $F(3, 463) = 3.03$ 1, $p = 0.02$, $\eta_p^2 = 0.019$. Follow-up analyses within site revealed that within the larger site, the interaction of pretest and treatment condition and main effect of treatment condition was not significant. Within the smaller site, the slope of posttest on pretest for students in large-group intervention was steeper than that of students in small-group intervention. No significant moderator or clustering effects were identified.

For the TOSRE, there was a three-way interaction of pretest, site, and age, $F(1, 449) = 4.02$, $p = 0.008$, $\eta_p^2 = 0.026$. To further examine the effects of age, the three treatment groups were plotted separately within site, with age dichotomized. The prediction was not different for older or younger students, with age not significantly influencing prediction. Regarding site, follow-up analyses were not significant, but the general pattern suggests that the slopes of posttest on pretest for students in the smaller site were steeper than students in the larger site. No significant moderator or clustering effects were identified.

For Passage Fluency, there was no overall difference in adjusted group means, $F(2, 453) = 2.70$, $p = 0.07$, $\eta_p^2 = 0.01$. There was an effect size of 0.10 favoring large-group treatment (difference between large-group treatment and business as usual) and 0.36 favoring small-group treatment (difference between small-group intervention and business as usual). No significant moderator or clustering effects were identified.

For WLF, there was a three-way interaction of pretest, treatment condition, and age, $F(2, 431) = 4.04$, $p = 0.0183$, $\eta_p^2 = 0.02$. To further examine the effects of age, the three treatment groups were plotted separately within site, with age dichotomized. The prediction was not different for older or younger students, with age not significantly influencing prediction. There was an effect size of 0.07 favoring the large-group intervention (difference between large-group treatment and business as usual) and 0.15 favoring the small-group intervention (difference between small-group intervention and business as usual). No significant moderator or clustering effects were identified.

Discussion

The purpose of this study was to examine the effectiveness of a reading intervention with older students with reading difficulties and the relative effects for two researcher-provided treatment groups that differed only on group size (not the type of treatment provided nor amount of treatment). Overall, findings revealed few statistically significant results or clinically significant gains associated with group size or treatment. We will discuss findings as they relate to the elements of

instruction provided and the associated outcome measures: decoding and spelling, comprehension, and fluency.

Group size

Findings revealed that students in the three conditions (large-group treatment, small-group treatment, and school comparison) did not demonstrate statistically significant main effects for decoding and spelling, although the pattern of the effects was in the direction of favoring small-group, researcher-provided intervention. The pattern of findings reveals that group size was not significantly powerful to consistently alter outcomes for older students with reading difficulties.

Outcome domain

There were also no statistically significant effects between the groups for comprehension outcomes. The only effect of note was for the higher percentage of students in the large-group treatment who passed the state-level reading test (76%) as compared with the percentage in the school instruction comparison group (57%) who passed the same test at the end of the year.

Several word reading fluency measures were associated with improved outcomes for small-group and large-group treatments over the comparison condition. The pattern of effects was in the expected direction, with both treatment groups outperforming school comparison and with small group outperforming large group. In all cases, the effects were small (0.07–0.29).

The overall outcome of this comprehensive, large-scale intervention is that few findings support claims that the treatment groups outperformed the school-provided instruction. In addition, group size—at least when variation is from 10 to 12 compared with 2 to 5—provided consistent statistically insignificant differences between conditions. While the pattern of effects is in the direction of favoring the small-group instruction over the large-group instruction and the treatment over school-provided instruction, the effects are small and infrequently statistically significant. What are the implications of these rather disappointing findings?

One interpretation is that the needs of these older readers with reading difficulties, many of whom in high-poverty settings, may be more extensive than can be addressed by providing one class (50 min) daily in reading instruction. Our clinical observations revealed that these students were significantly malnourished with respect to their understanding of word meanings, concepts, background knowledge, and critical thinking. Our intervention addressed these elements, but it may not have been adequate to meet these students' extensive needs. We were reminded daily of how these students understood few words, inadequately or incorrectly understood many concepts, and demonstrated inexperience in thinking critically about what they were reading. Many of these students were unfamiliar with identifying text they did not understand. They perceived of reading as "plowing through text" with little regard for what they thought or learned in the process. These observations suggest that a much more extensive, ongoing, and intensive intervention is required if students are to acquire the capacity to read and

learn from text. Such an intervention may be needed for multiple years and across multiple classes.

Our observations and reflections on the data lead us to consider whether any single intervention provided during one class period a day would adequately meet the extensive needs of these students. It may be tempting to suggest that the particular intervention we implemented was the wrong intervention and that a different intervention would yield more effective outcomes. We do not think that the extant research and our own clinical judgments support this view. For the target students represented in this study, based on clinical observation and findings from this and related studies, we believe that both earlier, extensive and ongoing interventions are required. Our observations also suggest that addressing the reading comprehension difficulties of these students is unlikely to result from a single intervention delivered solely by the reading or language arts teacher. School wide efforts that engage all content teachers in a unified approach to improving word and world knowledge are needed.

We appreciate the possible influence of contextual factors such as "choice" about taking the reading intervention (students in treatment condition) over taking an elective (students in comparison condition) and the possible influence that this might have on student engagement and performance. We have several data sources that suggest that students were at least as engaged during our intervention as they were during any other class period. First, none of the moderators were associated with outcomes, including site, fidelity, group size, or more intervention time. Second, fidelity measures reveal that students' engagement during instruction was high (see "Method"). Third, teachers' reported that students liked the class and found it to be one of the times during the day when they received the kind of personal attention and interest they craved. Fourth, students' attendance in the treatment condition did not differ from students' attendance in the comparison condition. At least based on the evidence we could obtain, we were unable to conclude that students' engagement was influenced negatively through participation in the treatment.

Connections with previous intervention studies with older students

Two large-scale intervention studies with older struggling readers relate directly to the findings from this study. The first large-scale study was conducted by the National Center for Education Evaluation and Regional Assistance (Corrin et al., 2008) and was aimed at improving the reading comprehension of struggling ninth-grade students who were descriptively similar to the students in this study with respect to ethnicity and pretest performance on reading comprehension measures. The intervention was provided daily as a supplement to their instruction and replaced an elective. The overall findings revealed no statistically significant differences between treatment groups and business as usual groups on vocabulary- or reading-related behaviors. Small differences (ES = 0.08) were reported for reading comprehension (Corrin et al., 2008); however, only one of the treatments was statistically significantly different from the comparison condition. This was the second year of the study and replicated the findings from the previous year (Kemple et al., 2008), in which the overall comprehension effect size was 0.09. A second

large scale study was conducted in Florida and explored the relative effectiveness of intensive reading interventions (provided for 90 min a day) to high-school students with reading difficulties on performance on the state reading assessment. For participating students reading below the fourth grade level, there were no statistically significant difference in any of the four treatment interventions provided. For students who were reading above fourth grade level, but still demonstrating reading difficulties, two of the four interventions produced significant gains for these students (Lang et al., 2009).

Similar findings were also reported in a study examining the effects of a reading intervention on sixth-graders with reading difficulties (Denton et al., 2008). These students were provided an intervention similar to the one provided to seventh- and eighth-graders in the study reported here; however, outcomes for the sixth-grade students yielded a larger effect on several reading measures, including passage comprehension (ES = 0.19). None of these intervention studies (Corrin et al., 2008; Denton et al., 2008; Kemple et al., 2008; Lang et al., 2009), including our study reported here, yielded results suggesting that an intervention provided over one school year was robustly effective, especially in terms of closing the gap between students with significant reading difficulties and their typically achieving peers.

Relevant considerations and limitations

This study was conducted in high-poverty schools with considerable daily challenges for the teachers and students. Findings from this study may have been considerably different if it were conducted in less challenging schools. It is likely that the contextual issues associated with high numbers of students from low socioeconomic status families makes issues related to schooling and literacy even more challenging.

The cost of any intervention is an important consideration. The small-group treatment intervention provided in this study was relatively costly with respect to typical class instruction, in which there may be one teacher for every 20–25 students. This cost needs to be considered with respect to the benefit. The practical benefit for the lives of the target students in this study could be considered minimal. It may be that practically significant differences for these students would require a more radical approach—either by increasing time or significantly altering the curricula throughout the school day. This study does not provide a specific future direction for what this type of intervention would be but does suggest that altering a single class is not likely to influence the quality of the reading comprehension of the majority of struggling readers in high-poverty schools.

This study also serves to remind us of the importance of prevention. Older students with reading difficulties demonstrate significant challenges for remediation and likely require multiple years of intervention. Prevention approaches that provide early intervention to students at risk and monitor their progress over time, continuing to provide intervention as needed, are an essential process for reducing the number of older at-risk readers (Fletcher, Lyon, Fuchs, & Barnes, 2006).

Our initial hypothesis was that small-group treatment would outperform both large-group treatment and the school-implemented comparison condition. Findings from this study suggest that educators consider models for response to intervention

(RTI) for older students that provide even more intensive interventions by increasing time or reducing group size. It may be that more intensive intervention requires very small groups or one-on-one instruction to realize the gains needed. Furthermore, it is reasonable to think that interventions that are more individualized and responsive to students' needs and less standardized might be associated with improved outcomes. These questions were not investigated in this study and would be valuable to address in future research.

References

Archer, A. L., Gleason, M. M., & Vachon, V. L. (2003). Decoding and fluency: Foundation skills for struggling older readers. *Learning Disabilities Quarterly, 26*, 89–101.

Archer, A. L., Gleason, M. M., & Vachon, V. (2005). *REWARDS intermediate: Multisyllabic word reading strategies*. Longmont, CO: Sopris West.

Biancarosa, G., & Snow, C. E. (2004). *Reading next—A vision for action and research in middle and high school literacy: A report from Carnegie Corporation of New York*. Washington, DC: Alliance for Excellence in Education.

Corrin, W., Somers, M., Kemple, J. J., Nelson, E., Sepanik, S., Salinger, T., et al. (2008). *The enhanced reading opportunities study: Findings from the second year of implementation*. Washington, DC: US Department of Education (NCEE 2009–4037).

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

Denton, C. A., Wexler, J., Vaughn, S., & Bryan, D. (2008). Intervention provided to linguistically diverse middle school students with severe reading difficulties. *Learning Disabilities Research and Practice, 23*(2), 79–89.

Dobson, A. J. (1990). *An introduction to generalized linear models*. London: Chapman and Hall.

Edmonds, M. S., Vaughn, S., Wexler, J., Reutebuch, C. K., Cable, A., Tackett, K., et al. (2009). A synthesis of reading interventions and effects on reading outcomes for older struggling readers. *Review of Educational Research, 79*(1), 262–300.

Fletcher, J. M., Lyon, G. R., Fuchs, L. S., & Barnes, M. A. (2006). *Learning disabilities: From identification to intervention*. New York: Guilford.

Green, P., & Silverman, B. (1994). *Nonparametric regression and generalized linear models*. Glasgow: Chapman and Hall.

Heward, W. L., & Silvestri, S. M. (2005). The neutralization of special education. In J. W. Jacobson, R. M. Foxx, & J. A. Mulick (Eds.), *Controversial therapies for developmental disabilities: Fad, fashion and science in professional practice* (pp. 193–214). Mahwah, NJ: Lawrence Erlbaum Associates.

Kamil, M. L., Borman, G. D., Dole, J., Kral, C. C., Salinger, T., & Torgesen, J. (2008). *Improving adolescent literacy: Effective classroom and intervention practices: A practice guide*. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, US Department of Education (NCEE 2008-4027).

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman brief intelligence test* (2nd ed.). Minneapolis, MN: Pearson Assessment (K-BIT-2).

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

Lang, L., Torgesen, J., Vogel, W., Chanter, C., Lefsky, E., & Petscher, Y. (2009). Exploring the relative effectiveness of reading interventions for high school students. *Journal of Research on Educational Effectiveness, 2*(2), 149–194.

Lee, J., Grigg, W. S., & Donahue, P. L. (2007). *The nation's report card: Reading 2007*. Washington, DC: National Center for Education Statistics, Institute of Education Sciences, US Department of Education (NCES 2007-4960).

Lexile Framework for Reading. (2007). Durham, NC: MetaMetrics.

McCray, A. D., Vaughn, S., & La Vonne, I. N. (2001). Not all students learn to read by third grade: Middle school students speak out about their reading disabilities. *Journal of Special Education, 35*, 17–30.

Muthen, L. K., & Muthen, B. O. (1998–2007). *MPlus statistical analysis with latent variables: User's guide* (5th ed.). Los Angeles: Muthen & Muthen (NAEP, 2007).

SAS Institute. (2002–2003). *SAS v. 9.1.3 Service Pack 4 [Computer program]*. Cary, NC: Author.

Scammacca, N., Roberts, G., Vaughn, S., Edmonds, M., Wexler, J., Reutebuch, C. K., et al. (2007). *Interventions for adolescent struggling readers: A meta-analysis with implications for practice*. Portsmouth, NH: RMC Research Corporation, Center on Instruction.

Shinn, M. R., & Shinn, M. M. (2002). *AIMSweb training workbook: Administration and scoring of reading maze for use in general outcome measurement*. Eden Prairie, MN: Edformation.

Snow, C. E. (2002). Second language learners and understanding the brain. In A. M. Galaburda & S. M. Kosslyn (Eds.), *Languages of the brain* (pp. 151–165). Cambridge, MA: Harvard University.

Texas Education Agency. (2004a). *Appendix 20—Technical digest 2004–2005*. Retrieved January 8, 2009, from http://ritter.tea.state.tx.us/student.assessment/resources/techdig05/appendices.html.

Texas Education Agency. (2004b). *Appendix 24—Technical digest 2004–2005*. Retrieved January 8, 2009, from http://ritter.tea.state.tx.us/student.assessment/resources/techdig05/appendices.html.

Texas Education Agency. (2004c). *Technical digest 2004–2005*. Retrieved January 8, 2009, from http://ritter.tea.state.tx.us/student.assessment/resources/techdig05/index.html.

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

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

Torgesen, J. K., Rashotte, C., Alexander, A. W., Alexander, J., & MacPhee, K. (2003). Progress toward understanding the instructional conditions necessary for remediating reading difficulties in older children. In B. Foorman (Ed.), *Preventing and remediating reading difficulties: Bringing science to scale* (pp. 275–297). Baltimore: York.

Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Rose, E., Lindamood, P., Conway, T., et al. (1999). Preventing reading failure in young children with phonological processing disabilities: Group and individual responses to instruction. *Journal of Educational Psychology, 91*, 579–593.

Vaughn, S., Fletcher, J. M., Francis, D. J., Denton, C. A., Wanzek, J., Wexler, J., et al. (2008). Response to intervention with older students with reading difficulties. *Learning and Individual Differences, 18*, 338–345.

Vaughn, S., Linan-Thompson, S., & Hickman, P. (2003). Response to instruction as a means of identifying students with reading/learning disabilities. *Exceptional Children, 69*, 391–409.

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., & Pearson, N. (In press). *The test of silent reading efficiency and comprehension (TOSREC)*. Austin, TX: PRO-Ed.

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

Wilder, A. A., & Williams, J. P. (2001). Students with severe learning disabilities can learn higher order comprehension skills. *Journal of Educational Psychology, 93*, 268–278.

Woodcock, R. W., McGrew, K., & Mather, N. (2001). *Woodcock-Johnson III tests of achievement*. Itasca, IL:Riverside.

Zeno, S. M., Ivens, S. H., Millard, R. T., & Duvvuri, R. (1995). *The educator's word frequency guide*. New York: Touchstone Applied Science Associates.