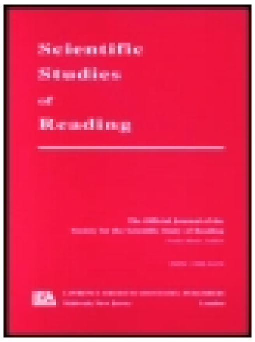

SCIENTIFIC STUDIES OF READING

ISSN: 1088-8438 (Print) 1532-799X (Online) Journal homepage: https://www.tandfonline.com/loi/hssr20

Supplemental Fluency Intervention and Determinants of Reading Outcomes

Patricia F. Vadasy & Elizabeth A. Sanders

SCIENTIFIC STUDIES OF READING, *13*(5), 383–425

ISSN: 1088-8438 print / 1532-799X online
DOI: 10.1080/10888430903162894

Supplemental Fluency Intervention and Determinants of Reading Outcomes

Patricia F. Vadasy and Elizabeth A. Sanders
Washington Research Institute, Seattle

This study replicates research on the efficacy of a repeated reading intervention with word-level instruction for students in Grades 2 and 3 with low to moderate fluency skills, examines differences between treatment implementers, and tests unique contributions of treatment-related variables on outcomes. Students from 13 schools were randomly assigned to dyads; dyads were randomly assigned to treatment or control conditions. Schools were matched into treatment implementer groups (teachers or paraeducators) at study onset. Tutoring occurred during school hours for 15 weeks (M = 25.5 hr). Multilevel model results showed treatment students (n = 98) gained more than controls (n = 104) on measures of letter-sound knowledge (d = .41), fluency (d = .37–.38), and comprehension (d = .30–.31); students tutored by teachers gained more than their paraeducator-tutored peers on word reading and fluency. Finally, dyads tutored with greater fidelity gained more in word reading and fluency; dyads that read more complex words in their texts gained less on letter-sounds, fluency, and comprehension.

Gough and Tunmer's (1986) "simple view" of reading (see also Gough, Hoover, & Peterson, 1996; Hoover & Gough, 1990) asserts that the development of two underlying language processes enables the reader to understand text. The simple view represents reading comprehension as the product of two independent components: decoding and language comprehension. As Adolf, Catts, and Little (2006) pointed out, the word recognition component of the simple view has typically been presumed to be word reading accuracy. There is considerable empirical support for the soundness of the simple view (Gough et al., 1996; Joshi & Aaron, 2000; Savage, 2001), and researchers have described developmental changes in the relative

Correspondence should be sent to Patricia F. Vadasy, Senior Researcher, Washington Research Institute, 150 Nickerson Street, Suite 305, Seattle, WA 98109. E-mail: pvadasy@wri-edu.org

contributions of the components to reading comprehension. Word recognition skills explain more variance in reading comprehension in the early stages of reading development, and in later grades the contribution of listening comprehension increases (Catts, Fey, Zhang, & Tomblin, 1999; Francis, Fletcher, Catts, & Tomblin, 2005; Gough et al., 1996; Rupley, Willson, & Nichols, 1998).

Most research on the simple view has focused on these two primary components and has not directly addressed the role of fluency, although others have proposed expanded models that include the unique role of reading speed (Cutting & Scarborough, 2006; Joshi & Aaron, 2000), and working memory (Seigneuric & Ehrlich, 2005). Adolf et al. (2006) directly addressed the contribution of a separate fluency component to the simple view in a series of concurrent and predictive structural equation models to examine variance in reading comprehension for a sample of students in second, fourth, and eighth grades. In their concurrent and prospective models, fluency did not account for unique variance in reading comprehension at any grade after controlling for word recognition and listening comprehension, and word reading accuracy accounted for nearly all variance in reading comprehension at the second-grade level. As Adolf et al. discussed, fluency is quite directly related to experience and skill in word recognition for novice readers and may be viewed as a product of lexical and sublexical efficiency that builds as word identification skills become accurate and automatic. These word-level skills then come to be applied in context reading, where word recognition and context facilitation occur in an interactive and mutually supportive manner (Stanovich, 1980). Fluency becomes an important instructional goal as children encounter longer texts that demand increasing allocation of cognitive resources to comprehend. Fluency skills are, not surprisingly, highly correlated with reading comprehension (Fuchs, Fuchs, & Maxwell, 1988; Jenkins, Fuchs, van den Broek, Espin, & Deno, 2003). Fluency deficits are therefore of concern because they represent an obstacle to comprehension and have been found difficult to remediate (Rashotte, MacPhee, & Torgesen, 2001; Torgesen, Rashotte, & Alexander, 2001; Vadasy & Sanders, 2008a, 2008b).

In this study, we evaluated the efficacy of a supplemental repeated reading intervention package that addresses fluency component skills particularly important in early reading development. We also consider features of instruction and assessment that are of broader theoretical and practical interest. These include the influence of implementer type in a fluency intervention, the student and intervention characteristics that influence fluency outcomes, and measurement issues regarding comprehension assessment.

Perfetti's (1977, 1985) verbal efficiency theory provides further background for this study. This theory addresses differences in reading comprehension that originate in lower levels of processing: orthographic, phonological, and semantic. Efficiency in these local processes, in particular accurate and rapid lexical access, enables the reader to allocate limited cognitive resources to comprehension processes. In the early stages of reading development, fluency reflects the consolidation of word-level processes, including alphabetic, decoding, and word identification skills (Kame'enui, Simmons, Good, & Harn, 2001; Rupley et al., 1998). Research underscores the important role of fluent word reading skill in the text reading fluency of younger and less skilled students (Jenkins et al., 2003; Schwanenflugel et al., 2006). This growing regard for the role of lower level component skills in emerging fluency informed the intervention approach we designed for the lower skilled second and third graders in this study.

STUDENT AND INTERVENTION CHARACTERISTICS THAT INFLUENCE FLUENCY

Repeated reading has strong support as a remedial approach to increase reading speed. First described by Dahl (1979) and Samuels (1979), repeated reading requires students to read a short passage repeatedly until reaching a criterion speed. A variety of repeated reading approaches have been tested, and Kuhn and Stahl (2003) reviewed 71 studies of assisted and unassisted repeated reading, classroom fluency approaches, and integrated fluency instruction that included book discussion. Overall, these approaches resulted in improved fluency and comprehension, when compared to regular classroom reading instruction. Less clear were the differential benefits of variations in fluency instruction, including difficulty of text, whether reading was assisted, and type and volume of reading material. A particularly comprehensive fluency intervention that included instruction and practice in lexical and sublexical as well as text-level fluency is described by Wolf, Miller, and Donnelly (2000). Their theoretically based small-group intervention for second and third graders at risk for reading disability addressed naming speed deficits that influence word recognition, vocabulary, and motivation. One component of the intervention emphasized phonological and orthographic pattern recognition hypothesized to support strong lexical representations in memory. Preliminary findings (Wolf & Katzir-Cohen, 2001) included significant gains in word attack, word identification, fluency, and comprehension.

In an earlier study (Vadasy & Sanders, 2008b), we examined the effects of the same *Quick Reads* (Hiebert, 2003) fluency intervention used in the present study, provided with similar incidental word-level scaffolding instruction, and implemented by paraeducator tutors. The second- and third-grade students in that study averaged at the 34th percentile in fluency at pretest and were randomly assigned to dyads and to treatment. Students received an average of 25 hr of instruction. In that study a series of classroom observations were conducted to quantify features of literacy instruction for all students in the study and to test the unique and conditional effects of classroom time allocated to oral reading on student gains. Treatment students made significantly greater gains than controls in word reading accuracy (d = .29) and fluency (d = .35), and pretest word reading uniquely predicted fluency rate gains. The amount of classroom oral reading time had small but reliable unique effects on gains in word reading accuracy (d = .30) and fluency (d = .35). In the present study, similar brief instruction and practice in phonological decoding and orthographic pattern recognition was added to a repeated reading intervention to address lexical and sublexical fluency components described in Schwanenflugel et al.'s (2006) *simple reading fluency model.* Their developmental model of fluency for early elementary children (Grades 1–3) suggests that word recognition efficiency may be the limiting factor in reading comprehension in early stages of reading development when word reading is not yet automatic. This is the stage, described as the "decoding reader stage" by Wolf (2007), in which the children in this study were operating.

One determinant of the effectiveness of many types of instruction is the skill level of the implementer. The literature on teacher quality indicates the importance of teacher verbal and cognitive ability in accounting for variance in student achievement (Ehrenberg & Brewer, 1994; Ferguson & Ladd, 1996; Greenwald, Hedges, & Lane, 1996; Kain & Singleton, 1996). Likewise, teacher experience has a significant effect on elementary student reading and math outcomes (Greenwald et al., 1996; Rowan, 2002). In this study we compare the effects of teacher- versus paraeducator-implemented supplemental instruction on student reading outcomes. Our understanding of the differential effects of teacher versus nonteacher instruction in reading skills currently requires extrapolation from studies that compare types of adult tutors (Pikulski, 1994; Wasik & Slavin, 1993) and employ nonexperimental designs. For example, in their meta-analysis of the effectiveness of one-to-one reading interventions for at-risk students, Elbaum, Vaughn, Hughes, and Moody (2000) found that across different types of reading interventions, children tutored by college students made the largest gains relative to teachers, community volunteers, and paraprofessionals. Grek, Mathes, and Torgesen (2003) conducted an observational study comparing the implementation of a highly prescriptive reading intervention by teachers and trained paraeducators. Teachers and paraeducators worked with small groups of first graders identified as at risk for reading failure and were observed by researchers over the course of the intervention. Observations captured teaching behaviors (e.g., pacing, adhering to lessons, correcting and scaffolding), as well as organization and management. Teachers were rated significantly higher in their instructional delivery, including pacing, error correction, maintaining student attention, and eliciting student responses. However, the researchers reported no differences between groups in student reading outcomes.

To our knowledge, the relative effectiveness of teacher versus paraeducator implementers has not been directly compared. This comparison would be inappropriate for complex multilayered interventions that would not be expected to be implemented without extensive training and experience but is particularly relevant in regards to a repeated reading intervention. Students who are not yet fluent readers are often held back in a negative feedback loop. They receive limited reading practice and experience low motivation for reading—both of which are essential for reading growth (Allington, 1977; Compton, Appleton, & Hosp, 2004; Stanovich, 1986). These students most need added word exposure and reading practice, the core features of most repeated reading interventions. A reading intervention to provide children with this added practice is one that might be implemented and effectively supported with minimal training. The practical importance of this comparison is underscored by the growing number of paraeducators in U.S. schools (Nelson, 1999), increasing more than 40% (in full-time equivalents) from 1990 to 1997. Yet these large numbers of school staff are often engaged in functions that have little bearing on student achievement (Gerber, Finn, Achilles, & Boyd-Zaharias, 2001) and are not adequately trained or matched with materials specifically designed or evaluated for use by paraeducator instructors. In this study we compare the effectiveness of paraeducators and teachers in providing a fluency intervention with word-level scaffolding for second and third graders with low to moderate word reading and fluency skills.

STUDENT GROUPING

Most research on student grouping strategies concerns classrooms or small groups and considers the effects of group ability composition on student outcomes. For example, meta-analyses of within-class grouping (Abrami, Lou, Chambers, Poulsen, & Spence, 2000; Lou, Abrami, & Spence, 2000; Lou et al., 1996) report advantages for small-group instruction when students are grouped for homogeneous ability and for students of higher ability. Several studies suggest that homogeneous pairings provide a more effective learning context for reasoning and complex cognitive tasks (Fuchs, Fuchs, Hamlett, & Karns, 1998; Phelps & Damon, 1989). Heterogeneous grouping often demands teacher skills in adjusting instruction and learning opportunities to meet the needs of diverse students and facilitate learning for students of varying abilities and skill levels.

In the present study, students were instructed in pairs for a highly prescribed instructional task that, unlike many peer-mediated interventions, provided limited opportunities for student interactions. The repeated reading routine was scripted and followed a daily sequence. The fluency intervention was particularly well suited for paraeducator implementation and easily adjusted to meet the needs of each student in the dyad. During each session, tutors reminded both students in each dyad to focus on their individual fluency goal, and although student interaction during the instruction was minimal, each dyad had its own dynamic that may have influenced student motivation. During the repeated readings, students had opportunities to model or correct a partner or assist a partner in decoding a word. When tutors conducted brief comprehension checks after a reading, or at the end of a passage, students had opportunities to listen to each other's responses, and the higher skilled student often provided a model for an accurate response. After each timed reading, tutors recorded each student's fluency rate. Tutors usually shared these rates with each student, and the slower readers in dyads may have been motivated to catch up with the stronger partner. Thus, although the dyad interactions in this study were constrained by the structure of this intervention, the dynamics of paired grouping were present but unmeasured in this study. Students were tutored in dyads rather than individually as a practical compromise for the study design. Although we believe that schools are most likely to implement this fluency intervention for small groups, assigning students to small groups for this study would have limited our ability to randomly assign students. On the other hand, employing individual tutoring for the study would have yielded less useful findings for schools.

COMPREHENSION MEASUREMENT

In measuring children's reading comprehension for the current study, we gave careful consideration to test format and validity issues that have been raised by others (see Fletcher, 2006). Two widely used response formats for assessing comprehension are the cloze task format and the multiple-choice format (questions that follow silent or oral passage reading). Responses on cloze task comprehension tests typically require children to pay attention to semantic and syntactic context cues in the immediate environment of the deleted word. For example, the item stem might be "The duck was swimming in the ___," and then the student is asked to supply a response. Correct responses would include, for example, "pond" or "water," whereas incorrect responses would include, for example, "street." As such, cloze task performance reflects sensitivity to sentence- or phrase-level constraints on meaning (Carlisle & Rice, 2004). In addition, researchers have found that word recognition and decoding skills account for more variance in cloze comprehension tests than in question-answering types of comprehension tests (Bowey, 1986; Cutting & Scarborough, 2006; Francis et al., 2005; Nation & Snowling, 1997; Spear-Swerling, 2004). In sum, cloze tasks are thought to tap "local" comprehension and require some decoding skill.

Comprehension tests involving passage reading provide information on students' ability to answer different types of questions about a passage. Nevertheless, a long history of research on multiple-choice comprehension tests has often raised the question of whether responses on multiple-choice tests primarily reflect test-taking strategies or true understanding (Anderson, Hiebert, Scott, &

Wilkinson, 1985; Farr, Pritchard, & Smitten, 1990). One approach to this validity question is to study passage independence (i.e., whether students can correctly answer a multiple-choice item without reading or listening to the passage to which the item pertains; Katz, Lautenschlager, Blackburn, & Harris, 1990; Pryczak, 1972). Most recently, Keenan and Betjemann (2006) investigated passage independence for multiple-choice items on one of the comprehension measures we used in the present study, the Gray Oral Reading Test (GORT; Wiederholt & Bryant, 2001). Using a modest sample ($n = 77$) of college students, they found that approximately 25% of the GORT items could be answered with 75% accuracy without having read the passages and that the accuracy across these "passageless" items averaged 57%. In the same study, a small subsample of children with reading disabilities ($n = 10$, M age = 11 years) were able to answer the same items with an average of 47% accuracy. Although this evidence is clearly troubling, it is unclear whether second- and third-graders (M age = 8 years) with low to moderate reading skill would be able to correctly guess as well as the college students and middle-school aged children used in Keenan and Betjemann's (2006) study.

Because comprehension was a major outcome of interest in the current study (along with fluency) we thought it necessary to use at least two (albeit imperfect) measures: the GORT and the Woodcock Reading Mastery Test. The GORT represents one widely used multiple-choice format that may reflect level of text interpretation, and the Woodcock represents a widely used cloze format that may tap more "local" comprehension. In addition, using the GORT permits a comparison of findings with other studies of repeated reading. In particular, Hiebert (2005) reported GORT outcomes in her study comparing fluency instruction using the *Quick Reads* passages to literature-based texts, and three other more recent studies examined the relation of fluency, comprehension, and word reading skills using GORT comprehension scores (Adolf et al., 2006; Cutting & Scarborough, 2006; Pierce, Katzir, Wolf, & Noam, 2007). On the other hand, like others (Keenan & Betjemann, 2006; Pierce et al., 2007), we are concerned that "the GORT may overestimate reading comprehension skill as many of the items can be answered correctly without reading the passage" (Pierce et al., 2007, p. 892). In fact, we employed an adapted scoring of the GORT to reflect only items pertaining to passages the students could read with some degree of fluency (i.e., the responses we used in the adapted scoring are those that students would be less likely to guess compared with passages the student could not read fluently).

Finally, it is also important to note that we used random assignment to treatment and control groups. As such, any guessing that may have occurred is equally likely to occur in either group. If anything, item issues with the GORT contribute more to unexplained variance (and perhaps inflated scores) than to biased inferences about group differences.

RESEARCH QUESTIONS

We have four research questions for the present study:

1. Can we replicate previous research (Vadasy & Sanders, 2008b) that demonstrated a modest causal link between repeated reading intervention and increased word reading and fluency skills for second and third graders?
2. Is there an advantage for students to be tutored by certificated teachers compared with paraeducators in word reading, fluency, and comprehension gains?
3. Is intervention fidelity uniquely related to individual differences in word reading, fluency, and comprehension gains?
4. Finally, are intervention text characteristics uniquely related to individual differences in word reading, fluency, and comprehension gains?

METHOD

Students

Referral and screening. In the fall of the academic year, second- and third-grade teachers from 13 urban, public elementary schools were asked to refer students who (a) had never been retained, (b) had low rates of reading fluency or comprehension, and (c) would benefit from a fluency intervention (i.e., students with adequate word reading skills to benefit from fluency instruction). Once active parent consents were obtained, referred students were screened for eligibility using two grade-level passages from the Oral Reading Fluency subtest of the Dynamic Indicators of Basic Early Literacy Skills (DIBELS; Good & Kaminski, 2002). Students whose performance was in the 10th to 60th percentile range on the average of the two passages were considered eligible for study participation (Hasbrouck & Tindal, 2006; 60th percentile was approximated by solving for the respective fluency rate value corresponding to $Z = 0.25$). This performance range was specifically 11 to 62 words correct per minute (wcpm) for second graders and 21 to 82 wcpm for third graders. We note here for ease of comparison that the same selection procedures were used in our prior research (Vadasy & Sanders, 2008b).

Group assignment. Treatment-control group assignment was a two-stage process. First, eligible students were randomly assigned to dyads (pairs of students) within grade and school using Microsoft Excel spreadsheet software's random number generator (uniform distribution). For schools with uneven numbers of students within grades, we used random selection to identify singletons that were subsequently excluded from participation. Dyads were then randomly assigned within school to one of two conditions: treatment (supplemental fluency tutoring instruction) or control (no tutoring; classroom instruction only). Again, this procedure was identical to our prior research (Vadasy & Sanders, 2008b).

Treatment implementer assignment. Schools were divided into two treatment implementer types: teacher-implementers (TI) and paraeducator-implementers (PI). Random assignment of schools to implementer type was not feasible because three schools had already hired paraeducators prior to study onset and only one tutor was typically required in each building. As such, treatment implementer type was determined by matching schools based on schools' indicators of academic risk (percentages of students who had failed the state reading subtest) and demographic risk (i.e., percentages of minorities, English language learners, and students eligible for free or reduced lunch; risk information reported on the state Web site for the prior academic year). Specifically, each school's risk value was standardized (into z scores), and then for each school we formed a composite by averaging across the risk indicators. Six schools were designated as TI research sites and seven were designated as PI sites. The two groups of schools did not significantly differ on any academic or demographic risk indicator (all t-test p values $> .05$).

Attrition. After group assignment, the sample comprised 230 students: 116 in the treatment conditions, and 114 in the control condition. At the end of the study, 18 treatment (6 from TI schools and 12 from PI schools) and 10 control students were lost to attrition (total attrition rate of 12%). To ensure an operational definition of treatment (again, similar to our prior study), if any treatment dyad member moved from the school, the corresponding member was also removed from study participation. Thus, the final sample includes 98 treatment students (54 TI and 44 PI students) and 104 control students. There were no significant differences between groups on demographic characteristics (see Table 1; all chi-square p values $> .05$).

Tutors

Fourteen tutors (all but one school used one tutor) were recruited from school communities. In 10 of the 13 participating schools, tutors were assigned to schools based on the implementer type we had already designated for schools, then taking into account classroom and tutoring schedules. Most tutors were female (100% and 75% for teachers and paraeducators, respectively) and White (83% and 63% self-report for teachers and paraeducators, respectively), and they had a median age of 45 years. Tutors' educational levels, general tutoring experience, and experience working with second and third graders varied. Whereas paraeducator tutors' highest educational attainment ranged from a high school diploma ($n = 5$) to a mas-

TABLE 1
Student Characteristics

	Treatment: Teachers (TI) n = 54		*Treatment: Paras (PI) n = 44*		*Controls: No Treatment n = 104*		*TI vs. TP*	*Treatment vs. Control*
	N	*%*	*N*	*%*	*N*	*%*	$\chi^2(1)$	$\chi^2(1)$
Grade 2	34	63%	32	73%	66	63%	1.05	0.34
Male	34	63%	24	55%	61	59%	0.71	0.01
Minority	35	65%	31	70%	68	65%	0.35	0.09
Asian	5	9%	6	14%	10	10%	0.47	0.14
Black	11	20%	8	18%	23	22%	0.07	0.23
Hispanic	15	28%	13	30%	28	27%	0.04	0.07
Mixed/Other	4	7%	4	9%	7	7%	0.09	0.15
Title I	38	70%	32	73%	80	77%	0.07	0.80
Special Education	3	6%	3	7%	6	6%	0.07	0.01
English Language Learner	16	30%	13	30	29	28%	0.07	0.07

Note. Chi-square tests of independence used to compare categorical frequencies between groups. All *p*s > .05

ter's degree ($n = 1$), all teacher tutors had master's degrees and were K-12 certificated ($n = 6$). Teachers averaged 14.2 years of classroom teaching experience (range = 4–32 years). With respect to tutoring experience, teachers averaged 0.7 years of experience ($SD = 1.21$) and paraeducators averaged 1.0 year of experience ($SD = 1.31$). Only one tutor (a paraeducator) had also served as a tutor in an earlier repeated reading intervention study. Finally, paraeducators and teachers were similar in their experience working with second and third graders ($M = 2.2$ years, $SD = 2.99$ for teachers and $M = 2.3$ years, $SD = 3.37$ for paraeducators). The only significant difference detected between teachers and paraeducators on demographic characteristics was their level of education, $\chi^2(2, N = 14) = 10.50$, $p < .01$.

Treatment

Treatment students received supplemental tutoring in dyads for 30 min per day, 4 days per week, for 15 weeks (December–April) using a modified application of the *Quick Reads* (Hiebert, 2003) fluency program (described next). Students assigned to the control group received regular classroom instruction, whereas treatment students received tutoring in their respective pairs. Classroom teachers reported that most treatment students ($n = 70$, or 71%) missed some portion of classroom literacy instruction, although this was more the case for the PI group (82%) than the TI group (65%), $\chi^2(1, N = 98) = 4.22$, $p < .05$. This is likely a reflection of differences in schools' schedules, however, because treatment implementer groups were designated by school. For the remaining treatment students (29%), the *Quick Reads* intervention served as added reading instruction time: These students typically missed nonreading activities and instruction such as math, science, social studies, art, physical education, computers, and library.

Program implementation. Standard *Quick Reads* procedures were modified in this study to address the instructional needs of students and to standardize implementation by tutors. Because it would be difficult for less experienced tutors to choose an appropriate teaching extension to add to the basic repeated reading steps, we created an instructional package that resembles what is suggested in the *Quick Reads* manual. *Quick Reads* is recommended for classroom or small group use, either as part of the regular reading program or as a supplemental intervention. The teacher's manual recommends three readings for each passage:

- *First read*: The teacher activates background knowledge about the topic, and students read the passage aloud or silently.
- *Second read*: The teacher reads aloud with the students, setting a model for fluent reading, all reading aloud at the target rate of 1 min. The teacher asks students to "tell the one thing the author wants you to remember."
- *Third read*: The students read silently for 1 min, and when the time is up each student records the number of words read. The teacher and student review comprehension questions together.

The manual also outlines placement procedures for students with passage reading performance of (a) at least 90% accuracy *and* at least a rate of 50 words per minute (wpm), (b) less than 90% accuracy *or* a rate of less than 50 wpm, and (c) less than 90 percent accuracy *and* a rate of less than 50 wpm. We followed placement guidelines for option (b), which were to "use your knowledge of the student to determine the appropriate instructional routine." Finally, the teacher's manual suggests teaching extensions to develop vocabulary, word identification, and comprehension skills. We added a brief word-level instructional extension to meet the needs of those students with especially low word-reading skills who were referred for study participation, and we allowed tutors up to 5 min per session for this extension. (As we describe next, we also measured the amount of phonics/word study instruction provided in the classroom to better understand how the intervention coordinated with classroom instruction for low decoders in particular.) All instruction for this intervention was scripted to ensure that all tutors used the same procedures. (The teacher manual is written for classroom or small-group instruction and assumes that the teacher chooses the extension activities, coordinated with student need and other reading instruction.) To incorporate the type of instructional support that a classroom teacher might provide, we scripted instructions for a brief layer of alphabetics/word study and trained tutors in effective word reading scaffolding and error correction procedures. Each tutoring session had six steps, as follows.

1. *Word reading instruction*: For this study we added one extension activity that, although not specifically prescribed in the teacher's manual, was appropriate for those students with word reading skills up to 1 *SD* below the mean, and brief enough not to weaken the recommended intensity of the core *Quick Reads* instruction (recommended for 12- to 15-min sessions to cover one passage). At the beginning of each tutoring session, tutors reviewed letter-sound correspondences. Tutors used a card featuring single letters and two-letter spelling patterns (e.g., *ea*, *ai*, *wh*), and selected items to review, based on student needs (as identified by pretests). Students reviewed the letters by either pointing to and reciting the letter name(s), picture cue name, and letter sound or simply pointing and saying the letter sound. Tutors added this alphabetic practice for up to 5 min of each session, as needed. As part of the passage reading, tutors also provided supportive correction strategies for students' word miscues. Trainers modeled how to scaffold student word reading attempts with graduated levels of assistance. Tutors were trained to assist students to apply the alphabetic skills they had practiced in isolation with the letter cards to their decoding attempts. Correction strategies for student miscues included (a) remind the student to use the letter-sound card to retrieve a letter sound; (b) encourage the student to sound out the word, phoneme by phoneme, and then blend the sounds; (c) help the student to segment a multisyllable word and then put the parts together; and (d) supply an irregular nondecodable word and have the student reread the word.
2. *First passage reading*: Tutors introduced the main idea for the first passage and pointed out one or two difficult words for students to practice before passage reading. Students then took turns reading sentences in the passage and following along with their finger. Tutors often used a transparency to underline student errors and review these words after the reading.
3. *Second and third passage reading*: The tutor and students read the passage aloud together twice, with the tutor modeling smooth, accurate, and fluent reading.
4. *Fourth passage reading*: Each student completed a 1-min timed reading for which the tutor recorded the students' reading rate and accuracy.
5. *Comprehension*: Tutors and students discussed two comprehension questions, one multiple choice, and one open ended.
6. *Read new passage/reread previous passage(s)*: As time permitted, students reread the previous passage or began a new passage (following the first five steps just outlined). Students completed Steps 1 to 5 for at least two passages per session.

Intervention text characteristics. The *Quick Reads* texts used in this study are short nonfiction passages written for grade levels 2 to 5 (the more recent published version also includes Grade 6). Each grade level includes six topics based on national standards for science and social science. Five passages are written to develop knowledge about each topic. For example, Level B (second grade) science topics include "Water and Us," "Weather," and "Rocks," whereas social studies topics include "Maps," "Money," and "Jobs Around Us." An important attribute of the *Quick Reads* program is the attention paid to text features to build fluency and comprehension. The teacher's manual describes texts as being designed to minimize the number of unique words (i.e., words are repeated to build opportunities for sight word learning) while maximizing grade-level phonics and syllable patterns (98% of words are described in the manual as high frequency, and more than 75% of words in second-grade texts were rated by Hiebert, 2005, as occurring in the most frequent word zones based on Zeno, Ivens, Millard, and Duvvuri, 1995). These text characteristics are hypothesized to assist in developing underlying lexical accuracy and automaticity skills often overlooked in traditional repeated reading interventions (Hiebert, 2008).

Because our fourth research question considers the influence of intervention text characteristics on treatment students' outcomes, we (a) analyzed a small sample of the *Quick Reads* passages (the first passage of each *Quick Reads* book of each level, amounting to three passages per level) to estimate mean Flesch Reading Ease and Flesch-Kincaid Grade Level for each *Quick Reads* level (Levels A–E) and (b) constructed a database of all of the words occurring in the *Quick Reads* texts to summarize word properties. To ensure that all words were accounted for, we scanned each of the *Quick Reads* passages and then imported the words into a text file that could then be imported into a database (with words cross-referenced for their *Quick Reads* book and level association). After we constructed this text database, we recorded each word's observed frequency (within *Quick Reads* book and level), standard frequency index value (SFI; Zeno et al., 1995), and decoding complexity index (DCI) value.

The SFI was used to describe overall word frequency. It is a logarithmically based measure of word frequency (per million words, weighted for content area in which words occur). According to Zeno et al. (1995), SFI values range from 3.5 (occurring less than once per million words) to 88.3 (occurring 67,500 times per million words) in the corpora of words collected (p. 12). The higher the SFI value, the more frequent a word is. One way to interpret SFI values is to examine the percentile ranks by grade level: According to Zeno et al. (1995), 50% of second-grade texts contain words with average SFI values of 39.5 or higher, whereas 50% of fourth-grade texts contain words with average SFI values of 34.4 or higher (p. 14; no third-grade values are provided). As grade level increases, less frequent words are expected. Using total words found in the *Quick Reads* passages (unique words weighted by their observed frequency), we found the mean SFI value to be 69.0

across all *Quick Reads* texts (see Table 2). Using only unique words, the mean SFI value is lower, at 53.8. According to Zeno et al. (1995, p. 14), only 5% of second-grade texts contain words with average SFI values as frequent as 57.5 (or higher), and 10% of second-grade texts contain words with average SFI values as frequent as 53.7. The observed word frequencies we found within the *Quick Reads* texts are in alignment with these SFI values: Only 2% of the total words occur one time across all texts (i.e., are singletons), and less than one third of all unique words occur once.

TABLE 2
Quick Reads Text Characteristics

Characteristic	*Level A (Grade 2)*	*Level B (Grade 2)*	*Level C (Grade 3)*	*Level D (Grade 4)*	*Level E (Grade 5)*	*All Levels*
Passage-Level Characteristics						
Mean Flesch Reading Ease	86%	87%	81%	62%	74%	78%
Mean Flesch-Kincaid Grade Level	4.2	3.9	4.7	8.1	5.9	5.3
Word-Level Characteristics, Total Words	8263	9205	1000 9	10925	11789	50191
Mean Observed Word Frequency	7.7	7.7	7.2	5.8	5.7	13.0
Mean Standard Frequency Index (SFI)	69.8	69.9	69.3	68.2	68.1	69.0
Mean Decoding Complexity Index (DCI)	6.1	6.1	6.2	6.6	6.7	6.4
Singletons	4%	4%	4%	7%	7%	2%
Word-Level Characteristics, Unique Words	1077	1201	1393	1885	2076	3857
Mean Standard Frequency Index (SFI)	59.7	59.1	58.4	56.5	56.2	53.8
Mean Decoding Complexity Index (DCI)	6.6	6.6	7.0	7.6	7.8	7.7
DCI Decodable Categories (1-9)	82%	82%	80%	76%	74%	77%
DCI Nondecodable Categories (10-11)	18%	18%	20%	24%	26%	23%
Singletons	30%	31%	32%	43%	42%	31%

Note. Each *Quick Reads* level contains three books with 30, 1-page, expository passages each (for a total of 90 passages/level); Flesch Reading Ease and Flesch-Kincaid Grade Level rated by *Microsoft Word 2003* on the first passage of each book within *Quick Reads* level (for a total of 3 passages/level); Observed Word Frequency = number of times words are observed to appear in *Quick Reads* texts; Standard Frequency Index (SFI) = log-transformed frequency per million words in total corpus from Zeno, Ivens, Millard, & Duvvuri (1995); Decoding Complexity Index (DCI) = 11-point rating scale of decoding complexity (categories are mutually exclusive) adapted from Compton, Appleton, and Hosp (2004) and Menon and Hiebert (1999); Singleton = words observed to appear only once in *Quick Reads* texts.

The DCI was used to describe the phonetic decodability of the words in the texts. The DCI is an 11-point rating scale adapted from Compton et al. (2004) and Menon and Hiebert (1999). Rating categories range from 1 (simple, single-syllable word patterns) to 11 (nondecodable, multisyllabic word patterns; see Table 3). Words rated between 1 and 9 represent those words that may be considered decodable (with higher scores indicating increasingly complex vowel and syllable patterns), whereas the upper ratings (10 and 11) represent words that may be considered nondecodable. Two graduate students in education were responsible for rating the words. Using a 10% random selection of unique words (n = 389), we found interrater reliability (r = .92, p < .001), and exact interrater agreement (83%) to be acceptable. As shown in Table 2, using total words (unique words weighted by frequency), mean DCI values, based on all words in texts, range from 6.1 to 6.7 across *Quick Reads* levels, with M = 6.4 (SD = 3.48). Using only unique words, mean DCI values are slightly higher, ranging from 6.6 to 7.8, with M = 7.7 (SD = 2.82). This disparity between mean DCI values based on total and unique words suggests that simpler, less complex words are employed in the *Quick Reads* texts more often than not. In fact, these data show that 77% of all unique words found in *Quick Reads* texts may be considered decodable (Table 2).

Intervention placement and coverage. The *Quick Reads* passages are organized by grade levels (A = second grade through E = fifth grade). Each *Quick Reads* level has three books, and each book contains six content areas with 5 passages per content area, for a total of 90 science and social science passages per grade level. Dyads were initially placed into their grade level of the *Quick Reads* texts (i.e., at study onset, all second graders were placed into Level B and all third graders into Level C). For each session, tutors recorded attendance and which *Quick Reads* passage(s) were covered during the tutoring session. By the end of intervention, treatment students covered an average of 94 passages (SD = 15.2) and attended an average of 51 tutoring sessions (SD = 6.8), or 25.5 hr of intervention. After computing students' individual passage coverage per session, we found that the treatment group averaged 1.8 passages per session (SD = 0.22, range = 1.1–2.7).

Although all dyads were initially placed into the first book of either Level B or C based on their grade level, some placements were adjusted after intervention began and coaches visited each tutor-dyad grouping to confirm or adjust the initial placement. Further, some dyads covered more passages than others by the end of intervention (i.e., some dyads moved through the intervention at a faster pace). As a result, there were eight combinations of passage level coverage patterns (e.g., one *Quick Reads* Level-Book pattern covered by nearly half of all dyads was B-1, B-2, B-3, and C-1). To compute the observed text characteristics our sample of students encountered, we calculated the text characteristics for each pattern and applied them to dyad(s) associated with each pattern. What is apparent from these

TABLE 3
Decoding Complexity Index Categories

Category	*Word Pattern*	*Additional Description*	*Examples*
1	A, I	single syllable; 1-letter words; 2-letter words with initial consonants	a, I
	C-V		me, we, my, by, so
2	C-V-C	single syllable; 3-letter words and 2-letter words with initial vowels	man, cat, hot
	V-C		am, if, ox, up, of, or
3	C-C-V	single syllable; consonant digraphs; consonant blends; consonant doublings	she, why, cry
	V-C-C-[C]		ash, itch, off
	C-C-[C]-V-C		that, chat, scrap
	C-V-C-C-[C]		tell, buff, back, mash, catch
	C-C-[C]-V-C-C-[C]		crash, track, scratch
	C-[C]-V-C-C		mask, hand, bent, stump, shall
4	[C]-[C]-[C]-V-C-e	single syllable; silent-e; includes plurals/past tense; includes soft c, soft g, and -re	ate, bake(d), mile(s), strike(s)
	[C]-[C]-[C]-V-R/G/C-e		splice, age, store
5	C-[C]-V-V-[C]-[C]	single syllable; vowel digraphs; includes added e (not silent-e)	pain, say, teach, bread, meet, chief, oats, book, moon, low, blue, key, hair, year, queen, great, steak
	V-V-C-[C]		eat, each
	C-[C]-V-V-C-[C]-E		breathe, Maine

6	Diphthongs	single syllable; diphthongs	haul, hawk, few, they, oil, trout, cloud, cow, boy
7	C-[C]-V-r	single syllabe; single vowels and vowel digraphs that are r- or l-controlled	car, scar, far, for, quartz
	[C]-[C]-V-r-C		farm, start, art, arm
	[C]-V-V-r		door
	[C]-[C]-V-ll		all, fall, small
	C-[C]-V-l-C		told, child
	C-[C]-V-V-l-C		could, field, build
	[C]-[C]-[C]-V-η-C-[C]-[C]		ink, ring, tank, strength
8	Multi-syllabic, decodable, inflected	two syllables; root word rated 1 to 7 (above) is inflected: -ed, -es, -ies, -ing, -ly, -est	roses, brewed, walking, bushes
9	Multi-syllabic, decodable, non-inflected	two or more syllables; all syllables rated 1 to 8 (above); not inflections	bankcards
10	Nondecodable monosyllabic	single syllable; cannot be rated 1 to 8; includes words with wr-, kn-, -dge, -nce	two, thought, there, friend, buy, find, right, know, suit, earn, course
11	Nondecodable multi-syllabic	two or more syllables; any part of word is nondecodable (not rated 1 to 9); includes -tion, -sion	dinosaur, petulant, communication

Note. Adapted from Compton, Appleton, and Hosp (2004) and Menon and Hiebert (1999). C = consonant, V = vowel, η = guttural nasal.

data (Table 4) is that the text characteristics our treatment groups were exposed to largely reflect what was observed in our text analyses of Levels B and C (Table 2).

Tutor training and coaching. Tutors participated in one initial 4-hr training by project staff. Training included an overview of reading fluency development and the repeated reading method. Research staff then modeled use of *Quick Reads* materials and demonstrated procedures for adding instruction/ scaffolding in decoding. Tutors practiced the protocols during training and received immediate feedback. Following this training, coaches visited tutors biweekly to provide follow up training and modeling and to collect data on protocol fidelity. Researcher coaches were assigned to specific tutors, and conducted a minimum of six observations on each tutor (of which there were at

TABLE 4
Sample-Specific Passage Coverage in *Quick Reads* Texts

Characteristic	*Treatment Teachers (TI) n = 54*	*Treatment: Paras (PI) n = 44*	*Combined Treatments n = 98*
Text Covererage By Level (% of Students)			
Level A (Grade 2)	7%	14%	10%
Level B (Grade 2)	74%	73%	73%
Level C (Grade 3)	81%	82%	82%
Level D (Grade 4)	19%	27%	22%
Level E (Grade 5)	0%	0%	0%
Passage-Level Characteristics			
Mean Flesch Reading Ease	84%	83%	84%
Mean Flesch-Kincaid Grade Level	4.4	4.5	4.4
Word-Level Characteristics, Total Words	11895	12428	12134
Mean Observed Word Frequency	8.2	8.3	8.2
Mean Standard Frequency Index (SFI)	69.6	69.6	69.6
Mean Decoding Complexity Index (DCI)	6.2	6.2	6.2
Singletons	4%	4%	4%
Word-Level Characteristics, Unique Words	1453	1502	1475
Mean Standard Frequency Index (SFI)	58.3	58.1	58.2
Mean Decoding Complexity Index (DCI)	6.9	6.9	6.9
Singletons	29%	29%	29%

Note. Each *Quick Reads* level contains three books with 30, 1-page, expository passages each; Flesch Reading Ease and Flesch-Kincaid Grade Level rated by Microsoft Word 2003 on the first passage of each book within *Quick Reads* level; Observed Word Frequency = number of times words are observed to appear in *Quick Reads* texts; Standard Frequency Index (SFI) = log-transformed frequency per million words in total corpus from Zeno, Ivens, Millard, & Duvvuri (1995); Decoding Complexity Index (DCI) = 11-point rating scale of decoding complexity (categories are mutually exclusive) adapted from Compton, Appleton, and Hosp (2004) and Menon and Hiebert (1999); Singleton = words observed to appear only once in *Quick Reads* texts.

least two observations per dyad). Coaches met monthly to discuss tutoring implementation progress.

Intervention fidelity. To monitor treatment implementation fidelity, data were collected via observation forms on (a) tutors' adherence to scripted *Quick Reads* protocols and (b) tutor instructional behaviors. Tutors' fidelity to protocols was measured using a 5-point rating scale of 1 (*never*) to 5 (*always*) for each intervention step just described. Tutor instructional behaviors were measured using the same 5-point scale for six criteria that included "Maximizes time on instruction," "Quick pace/smooth transitions/minimal pauses," "Uses appropriate specific praise," "Materials are organized," "Maintains accurate attendance records," and "Provides appropriate error correction/scaffolding." For each observation, mean tutor ratings for protocols and behaviors were used (rather than individual criterion ratings).

Prior to onsite tutor observations, we established interobserver reliability among the four researcher-observers using four videotaped *Quick Reads* sessions (featuring nonstudy tutors and students). Each observer viewed and rated each taped session using the observation criteria. From these ratings, interobserver reliabilities were computed. Internal consistency (Cronbach's alpha was computed using videotapes as "subjects" and observer mean ratings as "items") for the four raters was .96 for tutoring protocol fidelity, and .88 for instructional behaviors. To establish reliability for the new word reading component (which was not featured in the *Quick Reads* videos), we used five different videotaped sessions of a similar letter-sound activity from an earlier kindergarten intervention. Internal consistency among observers for this activity was .97.

Across 258 observations (ranging from 8 to 33 and averaging 18 per tutor), adherence to protocols averaged $M = 4.5$ ($SD = 0.37$) and tutor instructional behaviors averaged $M = 4.6$ ($SD = 0.37$). Although we found no statistically significant differences between tutor implementers, teacher-tutors tended to have higher ratings on the instructional protocols (implementation of intervention steps; $M = 4.7$, $SD = 0.18$) when compared with paraeducator-tutors ($M = 4.4$, $SD = 0.41$), $F(1, 12) = 3.87$, $p = .07$. Similarly, there was a tendency for teacher-tutors to have stronger instructional behaviors ($M = 4.8$, $SD = 0.21$) than paraeducators ($M = 4.5$, $SD = 0.40$), $F(1, 12) = 3.78$, $p = .08$.

Classroom Literacy Activities

To describe classroom literacy instruction received by students in the study, in February of the study year we surveyed 55 classroom teachers who had students participating in the study. Only 1 teacher (with four students: $n = 2$ treatment, $n = 2$ control) declined survey participation. Teachers reported spending $M = 124.63$ min ($SD = 34.95$) on literacy instruction in general, during which $M = 44.19$ min ($SD = 17.50$) were spent on oral and silent text reading, $M = 33.62$ min ($SD =$

15.16) were spent on vocabulary and comprehension, $M = 22.41$ min ($SD = 12.77$) on writing, and $M = 14.47$ min ($SD = 14.47$) on word reading. When we compared the treatment and control groups using simple t tests, we found no significant differences between conditions on any subcategory or total ($p > .05$).

Student Assessments

Students were individually assessed by trained testers unaware of group assignment on receptive language and rapid automatized naming (RAN) as well as three reading skills that were hypothesized to be differentially affected by intervention: word-level skills, reading fluency, and passage reading comprehension. Norm-referenced standard scores were used when available. In the measure descriptions that follow, published reliabilities are provided, as well as reliabilities for our sample (internal consistencies are Cronbach's alpha).

1. *Receptive language* was measured at pretest only with the standardized Peabody Picture Vocabulary Test–IIIA (Dunn & Dunn, 1997). Students select a picture that best illustrates the meaning of an orally presented stimulus word. Testing is discontinued after the student misses 8 out of 12 items. Test–retest reliability reported in the test manual is .93 for 6- to 10-year-olds. For this sample, internal consistency was .97.
2. *RAN* was measured at pretest only using the Letter Naming subtest of the Rapid Automatized Naming/Rapid Automatized Stimulus tests (Wolf & Denckla, 2005). For this 50-item subtest, students are presented with a card that has five randomly sorted letters (*a*, *d*, *o*, *p*, *s*) repeated 10 times each and are asked to say the names of the letters as quickly as they can. The raw score is the total number of seconds the student uses to name all of the letters. Test–retest reliability reported in the test manual is .87 for elementary grades. Because students were expected to have high accuracy on this subtest, we also measured students on the Number Naming subtest (*2*, *4*, *6*, *7*, and *9* are used as items) to obtain concurrent validity. For our sample, the correlation between Letter and Number Naming subtests (in terms of standard scores) was .71.
3. Because we had incorporated a word-level skills layer of instruction in the fluency intervention package, we measured *word-level skills* three ways: single letter sounds, letter combination sounds, and word reading. *Single letter sounds* were measured by the number of sounds the student correctly produces by reading a card with 25 randomly sorted single letters of the alphabet, with 25 letters possible (all letters except *q*). Hard consonants and short vowels were considered correct. Internal consistencies for our sample were .72 and .65 at pretest and posttest, respectively; the lower relia-

bilities reflect the high negative skew in the performance distribution on this measure. *Letter combination sounds* were measured by the number of 2-letter spelling patterns the student identifies correctly by reading a card with 42 randomly sorted letter combinations. Letter combinations included *qu*, 15 vowel teams (e.g., *ai*), 5 *r*-controlled vowels (e.g., *ir*), 14 double-letter consonant blends (e.g., *st*), and seven digraphs (e.g., *sh*). Internal consistencies for our sample were .88 and .86 at pretest and posttest, respectively. *Word reading* was measured using the Word Identification subtest from the norm-referenced, standardized Woodcock Reading Mastery Test–Revised/Normative Update Form H (WRMTR/NU; Woodcock, 1987/1998). This assessment requires students to read increasingly difficult words, and testing is discontinued after six consecutive incorrect responses. Split-half reliability (alternating items) reported in the test manual averages .99 for first and third graders. Internal consistencies for our sample were .94 at pretest and .93 at posttest.

4. *Reading fluency* was assessed two ways: in a raw-score framework as *passage reading fluency* (PRF) and in a norm-referenced framework as *fluency rate*. Although a norm-referenced measure of fluency can provide a generalizable measurement of students' fluency rate, we considered PRF—in terms of wcpm performance—to be a meaningful indicator of students' responsiveness to intervention. Students were assessed on PRF at pretest, midtest, and posttest, but in the interest of brevity we analyze and report data from pretest and posttest only. Thus, we measured students' PRF using passages drawn from DIBELS Oral Reading Fluency Benchmarks: one grade-level *uniform* passage used at each test interval (*uniform passage* for second graders was "If I had a Robot," and for third graders was "My Friend"), and one grade-level *alternate* passage at each test interval. Alternate passages for second graders were "Roller Coaster" and "Drift Bottle" at pretest and posttest, respectively; for third graders, alternate pretest and posttest passages were "Fieldtrip" and "Parents," respectively. For each uniform (PRF-U) and alternate (PRF-A) passage, students read aloud while the tester recorded errors, and testing was discontinued after 1 min. Words omitted, substituted, and hesitations of more than 3 sec are scored as errors (words self-corrected within 3 sec are scored correct). For our sample, the correlations between the uniform second- and third-grade passages were .88 and .90 at pretest and posttest, respectively, and the correlations between the uniform and alternate passages across both grade levels were .83 and .84 for pretest and posttest, respectively.

We assessed PRF using both uniform (PRF-U) and alternate (PRF-A) passages to minimize (a) measurement error from potential passage nonequivalence effects (by using a uniform passage) and (b) passage memory effects (by using alternate passages). Although the use of a uniform passage introduces potential memory effects (i.e., the student may remember words in a passage read on a previous occasion or the student may remember the story structure) and lower generalizability (i.e., the uniform passage has a specific vocabulary and story structure that differs from other passages), any memory effects occurring for treatment students have an equal chance of occurring for control students. Whereas mean estimates of true gains across all students on the uniform passage are likely to be inflated, estimates of *treatment effects* on gains using the uniform passage should be free of passage nonequivalence effects. Consistent with our reasoning, Jenkins, Zumeta, Dupree, and Johnson (2005) found that uniform passages (compared to alternate passages) measured fluency gains more reliably; however, memory effects were detected for the uniform passage at the 5-week retest.

Our second, norm-referenced assessment framework for measuring fluency (*fluency rate*) employed the Rate subtest from the GORT–4 Form B (Wiederholt & Bryant, 2001). For this subtest, students read passages aloud, beginning at their grade level. Performance on the grade-level passage determined whether students subsequently read a lower or higher level passage, and testing was discontinued when a basal and ceiling were established. The amount of time it took students to read each passage converts to a raw score, ranging from 0 to 5 for each passage read. To be consistent with our other norm-referenced measures, we transformed the GORT standard score distribution with a mean of 10 and standard deviation of 3 to have a mean of 100 and standard deviation of 15. Internal consistency reported in the test manual is .92 for 7- to 9-year-olds. Sample internal consistencies were .82 and .88 at pretest and posttest, respectively.

5. *Comprehension* was measured two ways: using a *cloze-task* measure (the WRMT–R/NU Passage Comprehension subtest) and using a *multiple-choice* measure (GORT–4 Comprehension subtest). The WRMT–R/NU Passage Comprehension subtest (our cloze-task measure) requires students to supply a missing word that would be appropriate in the context of each passage silently read. A series of acceptable responses are listed on the easel page for the tester. Items are increasingly difficult, and testing is discontinued after six consecutive incorrect responses. The test manual reports split-half reliability (alternating items) as averaging .97 for first and third graders. Sample internal consistencies were .90 at pretest and .87 at posttest.

   Our multiple-choice comprehension task was measured with the GORT–4 Comprehension subtest. Once the student reads a passage aloud (see our previous *fluency rate* description), the passage is removed from view and the student is provided with a copy of five comprehension questions along with each question's corresponding four choices. The tester reads each question and corresponding choices aloud for the student and then asks the student to choose the correct answer. The tester begins at a grade-level passage and then continues testing on other passages until the student misses three out of five of the comprehension questions for a given passage (the ceiling). The basal is established when the student correctly answers all five questions for a given passage. Internal consistency reported in the test manual is .96 for 7- to 9-year-olds. For our sample, internal consistencies were .86 and .89 at pretest and posttest, respectively.

It is important to note that because the GORT Comprehension subtest basal and ceiling scores are computed separately from the Rate subtest basal and ceiling, students may be able to answer comprehension questions for passages that they are not able to read (fluently or at all). Therefore, we applied a second raw scoring method (fluency-based) and scored only items *from those passages that were administered within the Rate subtest's basals and ceilings*. With this score we were able to measure comprehension for passages that students actually read with some degree of fluency. For both scoring methods, raw scores were transformed to standard scores.

Data Analysis Methods

We adopted multilevel modeling as our primary analytic method (also known as hierarchical linear modeling and random effects modeling) for testing our research questions. As compared with unilevel statistical methods (such as fixed-factor analyses of variance or multiple linear regression), the more complex multilevel analysis method accounts for dependencies in the data, allowing for valid inferences to be drawn about relationships between outcomes and predictors without violating the assumption of independence.

Research Questions 1 and 2 (treatment efficacy and implementer comparison). Because dyads were our unit of random assignment, and students were assigned to dyads across classrooms (rather than within classrooms), we employed cross-classified multilevel modeling to test treatment effects on outcomes (Research Questions 1 and 2). Cross-classified models are similar to two-level models: students (Level 1) nested within both dyads and classrooms (Level 2). These models account for variance between dyads as well as variance between classrooms (or in other words, account for the within-dyad and within-classroom correlations in the data). Variance between schools was ignored (although some part of this variance will be accounted for by classrooms and dyads), as three-level cross-classified modeling is not currently feasible.

To test our first two research questions, our three groups were contrast-coded into two vectors constructed to simultaneously test the effects of (a) any tutoring compared with no tutoring (dyads in both treatment groups coded with a +1 and dyads in the control group coded with –2 in the first vector) and (b) teacher tutoring compared with paraeducator tutoring (in the second vector, dyads with teacher implementers coded with +1, dyads with paraeducator implementers coded with –1, and dyads in the controls coded with 0 in the second vector). The cross-classified model we used for testing treatment effects on pretest–posttest gains is as follows (cf. Raudenbush, Bryk, Cheong, & Congdon, 2004, pp. 191–192).

$$Y_{ijk} = \theta_{000} + \beta_{01}*(\text{Slope T v C})_j + \beta_{02}*(\text{Slope TI v PI})_j + b_{00j} + c_{00k} + e_{ijk}.$$

In this model, each score Y (whether it be pretest or a pretest–posttest gain score) for the ith student in the j-kth dyad-classroom cell is equal to (a) the intercept θ_{000} (grand mean score, holding all other predictors constant at zero), plus (b) β_{01}, the fixed effect of experimental condition (+1 = any tutoring, –2 = no tutoring), plus (c) β_{02}, the fixed effect of tutoring implementer type (+1 = teacher implementer, –1 = paraeducator implementer, 0 = control), plus (d) b_{00j}, the residual between the student's score and their respective mean dyad score, plus (e) c_{00k}, the residual between the student's score and their respective mean classroom score, plus (f) e_{ijk}, the within-cell residual.

Research Questions 3 and 4 (unique effects of intervention fidelity and text characteristics). To test the effects of treatment-related variables on treatment students' outcomes (Research Questions 3 and 4), we used standard two-level modeling (students nested in dyads) rather than cross-classified modeling. This was because there were no degrees of freedom available to estimate the residual within-cell variance when we only include treatment students in the analyses. Specifically, the degree of freedom for the residual variance estimation must be positive and is calculated as

$$T - K*(\text{random effects for rows}) - Q*(\text{random effects for columns}),$$

where T = number of Level 1 units (students), K = number of row units (dyads), and Q = number of column units (classrooms) (M. Dutoit, personal communication, July 18, 2008). If we had attempted to estimate a cross-classified model for testing our third research question, our degree of freedom would be less than 1 (98 students – 49 dyads – 50 classrooms), which is not permissible. Thus, we opted to use a basic two-level model to answer the third research question as follows.

$$Y_{ij} = \gamma_{00} + \gamma_{10}*(\text{Pretest})_{ij} + \gamma_{01}*(\text{Tutoring Fidelity})_j + \gamma_{02}*(\text{Text Characteristic})_j + U_{0j} + r_{ij}.$$

In this model, a gain score for *i*th student in the *j*th dyad is equal to (a) γ_{00}, the intercept (grand mean gain, holding all other predictors constant), plus (b) γ_{10}, the fixed effect of Pretest (a student-level predictor that is grand-mean centered across all students), plus (c) γ_{01}, the fixed effect of Tutoring Fidelity (a dyad-level predictor that is grand-mean centered across all dyads), plus (d) γ_{02}, the effect of Text Characteristic (also a dyad-level predictor that is grand-mean centered across all dyads), plus (e) U_{0j}, the residual between the student's score and their respective mean dyad score, plus (f) r_{ij}, the within-dyad residual.

For multilevel analyses, *HLM 6.0* was used (Raudenbush & Bryk, 2002; Raudenbush et al., 2004); for all other analyses, SPSS 13.0 and Microsoft Excel 2003 were used.

RESULTS

Pretests

Observed group means and standard deviations across all measures are shown in Table 5. The pretest means for norm-referenced measures show that the sample averaged in the lower 30th percentile in receptive language, lower 25th percentile in RAN, the lower 45th percentile in word reading, the lower 10th percentile in fluency rate, and lower 20th and 30th percentile in comprehension for the cloze task and multiple-choice tasks (both scoring methods), respectively. When we applied the Hasbrouck and Tindal (2006) grade-based norms to students' performance on the uniform (PRF-U) and alternate (PRF-A) PRF measures, we found that the sample averaged between the 25th and 50th percentiles at pretest. Finally, we found no significant differences between treatment and control groups or between treatment implementer conditions (TI andPI) on pretest cross-classified models (Table 6).

Research Question 1: Treatment Efficacy on Pretest–Posttest Gains

Prior to testing group differences on gains, we estimated the intraclass correlations (ICCs) associated with classrooms (or, in other words, the percentage of variance in gains accounted for by classroom membership) as well as the ICCs associated with dyads (separately), for each measure, irrespective of experimental condition (often these models are called two-level random intercept models). Using restricted maximum likelihood estimation, ICCs associated with classrooms ranged from .08 (PRF-A gain) to .32 (PRF-U gain), and averaged .16 (i.e., on average, 16% of the variance in gains is estimated to be accounted for by classroom membership). With only one exception (PRF-A), significant between-classroom variance was detected on all measures of gains (chi-square p values < .05). The ICCs

TABLE 5
Observed Student Assessment Means and Standard Deviations

	Treatment: Teachers (TI) n = 54						*Treatment: Paras (PI) n = 44*						*Controls: No Treatment n = 104*					
	Pretest		*Posttest*		*Gain*		*Pretest*		*Posttest*		*Gain*		*Pretest*		*Posttest*		*Gain*	
Measures	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Receptive Language	90.46	18.30					92.27	16.46					90.81	18.89				
Rapid Naming	88.04	9.36					92.11	11.20					88.13	12.25				
Word-Level Skills																		
Single Letter Sounds	23.87	1.73	24.83	1.81	0.96	2.07	23.57	2.28	24.45	2.02	0.89	1.97	23.64	2.75	24.45	1.76	0.81	2.78
Letter Combo Sounds	26.19	7.32	34.48	6.35	8.30	6.49	23.11	8.31	31.86	5.49	8.75	6.53	25.03	7.15	30.90	5.90	5.88	5.79
Word Reading	96.06	7.38	99.91	8.46	3.85	7.30	96.89	8.34	97.70	6.32	0.82	4.52	96.75	8.71	98.66	9.24	1.91	6.23
Reading Fluency																		
PRF-U	42.26	19.10	90.09	20.30	47.83	22.75	43.64	18.00	88.39	20.97	44.75	19.62	44.50	19.59	83.28	21.73	38.78	17.75
PRF-A	39.37	16.60	77.98	21.93	38.61	22.16	41.36	14.20	70.86	27.45	29.50	19.97	40.18	14.33	70.30	24.48	30.12	17.56
Fluency Rate	76.02	8.38	90.00	12.92	13.98	11.83	78.52	8.73	90.68	13.06	12.16	11.12	78.13	8.87	87.45	13.11	9.33	9.63
Comprehension																		
Cloze-Task	92.63	6.77	96.33	9.02	3.70	7.65	93.23	8.01	95.34	7.07	2.11	5.25	92.94	7.68	93.97	7.68	1.02	5.18
Mult Choice	84.91	12.34	93.70	15.88	8.80	13.10	88.75	12.90	94.89	16.51	6.14	14.14	84.57	11.72	92.21	14.60	7.64	14.76
Mult Choice, FB	83.52	8.94	93.33	10.73	9.81	9.11	85.68	8.67	93.75	11.47	8.07	9.66	84.95	8.46	91.11	9.70	6.15	9.14

Note. Receptive Language = standard score of Peabody Picture Vocabulary Test Version 3; Rapid Naming = standard score of Letter Naming subtest from Rapid Automatized Naming and Stimulus Subtests; Single Letter Sounds = number of correctly read sounds of isolated letters (25 possible); Letter Combo Sounds = number of correctly read sounds of letter combinations (42 possible); Word Reading = standard score of Word Identification subtest from Woodcock Reading Mastery Test-Revised/Normative Update; PRF = passage reading fluency as measured by uniform (U) and alternate (A) passages drawn from the Dynamic Indicators of Basic Early Literacy Skills ORF benchmarks; Fluency Rate = standard score of Rate subtest from Gray Oral Reading Test Version 4 (GORT-4); Cloze-Task = standard score of WRMT-R/NU Passage Comprehension subtest; Mult Choice = standard score of Comprehension subtest from GORT-4; Mult Choice, FB = adapted, fluency-based standard score of Comprehension subtest from GORT-4.

TABLE 6
Pretest Cross-Classified Model Results

	Fixed Effects									Random Effects						
	Intercept (Pretest)			Slope (T v C)			Slope (TI v PI)			Classrooms			Dyads			Residual
	M	*SE*	*t(199)*	*M*	*SE*	*t(199)*	*M*	*SE*	*t(199)*	*Var*	χ^2 *(52)*		*Var*	χ^2 *(102)*	%	*Var*
Receptive Language	92.49	2.06	44.82***	0.07	0.69	0.10	0.96	1.84	0.52	159.01	213.04***	.47	9.08	113.39	.03	172.58
Rapid Naming	89.42	0.87	102.94***	0.66	0.53	1.24	–2.03	1.16	–1.75	1.41	52.29	.01	2.73	110.63	.02	122.02
Word-Level Skills																
Single Letter Sounds	23.65	0.22	109.43***	–0.01	0.11	–0.09	0.33	0.26	1.28	0.87	90.53***	.15	0.01	96.75	.00	4.81
Letter Combo Sounds	25.00	0.73	34.33***	–0.06	0.31	–0.18	0.62	0.78	0.79	14.41	132.57***	.27	0.07	93.82	.00	39.63
Word Reading	96.56	0.64	150.83***	–0.08	0.40	–0.21	–0.45	0.86	–0.52	0.45	51.73	.01	3.31	114.39	.05	63.98
Reading Fluency																
PRF-U	44.05	1.94	22.71***	–0.42	0.83	–0.50	–0.82	2.10	–0.39	99.84	120.31***	.28	30.12	128.08*	.08	231.89
PRF-A	40.37	1.31	30.75***	0.11	0.75	0.14	–0.85	1.70	–0.50	12.66	68.85	.06	47.98	157.74***	.22	157.06
Fluency Rate	77.54	0.68	114.50***	–0.27	0.42	–0.64	–1.25	0.92	–1.37	0.02	49.25	.00	6.44	124.32	.09	68.31
Comprehension																
Cloze-Task	92.93	0.57	163.75***	–0.01	0.36	–0.02	–0.30	0.77	–0.39	0.05	51.76	.00	0.68	106.75	.01	55.30
Mult Choice	86.09	1.02	84.16***	0.78	0.57	1.38	–1.71	1.30	–1.32	10.88	75.85*	.15	5.70	111.01	.08	55.34
Mult Choice, FB	84.65	0.75	112.84***	–0.10	0.40	–0.26	–1.03	0.94	–1.09	7.01	77.34*	.10	4.97	118.03	.07	61.28

Note. Receptive Language = standard score of Peabody Picture Vocabulary Test Version 3; Rapid Naming = standard score of Letter Naming subtest from Rapid Automatized Naming and Stimulus Subtests; Single Letter Sounds = number of correctly read sounds of isolated letters (25 possible); Letter Combo Sounds = number of correctly read sounds of letter combinations (42 possible); Word Reading = standard score of Word Identification subtest from Woodcock Reading Mastery Test-Revised/Normative Update; PRF = passage reading fluency as measured by uniform (U) and alternate (A) passages drawn from the Dynamic Indicators of Basic Early Literacy Skills ORF benchmarks; Fluency Rate = standard score of Rate subtest from Gray Oral Reading Test Version 4 (GORT-4); Cloze-Task = standard score of WRMT-R/NU Passage Comprehension subtest; Mult Choice = standard score of Comprehension subtest from GORT-4; Mult Choice, FB = adapted, fluency-based standard score of Comprehension subtest from GORT-4.

* $p < .05$. **$p < .01$. ***$p < .001$.

associated with dyads ranged from .001 (cloze task comprehension) to .27 (combination letter sounds), and averaged .10 (i.e., on average, 10% of the variance in gains is estimated to be accounted for by dyad membership). Significant between-dyad variation was detected for two outcomes: combination letter sounds and PRF-U (chi-square p values $< .05$).

Next, we tested for differences between groups on pretest–posttest gains using a series of cross-classified models (to account for variation between dyads and classrooms simultaneously). Results showed that the combined treatment group gained significantly more than controls on one of three word-level measures (combination letter sounds), two of three fluency measures (PRF-U and fluency rate), and two of three comprehension measures (cloze task and the adapted fluency-based scoring for our multiple-choice task; Table 7).

Specifically, holding differences between the two tutored groups constant (and keeping in mind the contrast coding we used), the combined treatment group gained an estimated 8.36 combination letter sounds correct (7.52 + 0.84*(+1)), whereas controls gained an estimated average of 5.84 sounds correct (7.52 + 0.84*(–2)), $t(199) = 2.99$, $p < .01$, $d = .41$. All of the other treatment effects detected are similarly interpreted. For fluency, treatment students gained an estimated 45.85 wcpm on PRF-U (uniform passage reading fluency measure) compared with an estimated 38.62 wcpm gained by controls, $t(199) = 3.07$, $p < .01$, $d = .37$; further, treatment students gained an estimated 13.28 standard score points on fluency rate (GORT Fluency subtest) compared with an estimated 9.35 standard score point gain estimated for controls, $t(199) = 2.83$, $p < .01$, $d = .38$. For comprehension, treatment students gained an estimated 2.94 standard score points on the cloze-task measure (WRMT-R/NU Passage Comprehension subtest) compared with controls who gained an estimated 1.08 standard score points, $t(199) = 2.27$, $p < .05$, $d = .31$; likewise, the treatment group gained an estimated 9.19 standard score points on the fluency-based multiple choice measure (adapted scoring for the GORT Comprehension subtest) compared with controls who were estimated to have gained 6.40 standard score points, $t(199) = 2.17$, $p < .05$, $d = .30$. The average treatment effect size (for significant effects) is $d = .35$, which is considered low to moderate by Cohen's (1988) standards. Nevertheless it is noteworthy that others, such as Rosenthal and Rosnow (1984), have pointed out that Cohen's standards underestimate effect sizes for variables measured with substantial error, such as the passage reading fluency measures used in this study.

Research Question 2: Treatment Implementer Comparison on Pretest–Posttest Gains

In addition to testing treatment efficacy (i.e., the combined treatment group compared with controls), our cross-classified models also simultaneously tested differences between students tutored by teachers (TI) and those tutored by paraedu-

TABLE 7

Pretest–Posttest Gain Cross-Classified Model Results

	Fixed Effects											Random Effects						
	Intercept (Gain)			Slope (T v C)				Slope (TI v PI)				Classrooms			Dyads			Residual
Measure	*M*	*SE*	*t(199)*	*M*	*SE*	*t(199)*	*d*	*M*	*SE*	*t(199)*	*d*	*Var*	$\chi^2(52)$	%	*Var*	$\chi^2(102)$	%	*Var*
Word-Level Skills																		
Single Letter Sounds	0.92	0.24	3.89***	0.05	0.10	0.53	.07	–0.30	0.26	–1.18	–.25	1.41	120.61***	.24	0.01	92.78	.00	4.45
Letter Combo Sounds	7.52	0.54	13.96***	0.84	0.28	2.99**	.41	0.06	0.66	0.09	.02	4.27	75.10*	.12	1.46	112.13	.04	31.28
Word Reading	2.31	0.57	4.09***	0.07	0.28	0.26	.03	1.34	0.67	2.01*	.43	6.15	96.39***	.16	0.05	90.41	.00	32.42
Reading Fluency																		252.598
PRF-U	43.44	2.00	21.73***	2.41	0.79	3.07**	.37	0.66	2.03	0.33	.07	123.58	154.03***	.33	0.48	91.70	.00	341.01
PRF-A	32.85	1.57	20.88***	1.40	0.91	1.53	.22	4.17	2.04	2.04*	.43	16.97	64.79	.05	12.91	112.12	.03	90.70
Fluency Rate	11.97	0.96	12.46***	1.31	0.46	2.83**	.38	0.46	1.12	0.41	.09	18.49	95.31***	.17	0.34	100.41	.00	
Comprehension																		
Cloze-Task	2.32	0.50	4.65***	0.62	0.27	2.27*	.31	0.72	0.63	1.15	.24	2.84	76.79**	.08	0.04	85.89	.00	32.19
Mult Choice	7.63	1.28	5.98***	–0.13	0.71	–0.18	–.03	1.52	1.63	0.93	.22	15.86	75.27	.08	41.99	159.83***	.21	140.12
Mult Choice, FB	8.26	0.85	9.75***	0.93	0.43	2.17*	.30	0.91	1.02	0.90	.20	12.30	91.20***	.15	6.69	121.91	.08	65.13

Note. Pretest–posttest gains analyzed. Contrast Coding used: T v C = both Treatment groups coded +1, Control group coded -2; TI v PI = Teacher-Implementer Treatment group coded +1, Paraeducator-Implementer group coded -1; Control group coded 0. Single Letter Sounds = number of correctly read sounds of isolated letters (25 possible); Letter Combo Sounds = number of correctly read sounds of letter combinations (42 possible); Word Reading = standard score of Word Identification subtest from Woodcock Reading Mastery Test-Revised/Normative Update; PRF = passage reading fluency as measured by uniform (U) and alternate (A) passages drawn from the Dynamic Indicators of Basic Early Literacy Skills ORF benchmarks; Fluency Rate = standard score of Rate subtest from Gray Oral Reading Test Version 4 (GORT-4); Cloze-Task = standard score of WRMT-R/NU Passage Comprehension subtest; Mult Choice = standard score of Comprehension subtest from GORT-4; Mult Choice, FB = adapted, fluency-based standard score of Comprehension subtest from GORT-4.

* $p < .05$. **$p < .01$. ***$p < .001$.

cators (PI). Results showed that the TI group gained significantly more than the PI group on one word-level measure (word reading) and one fluency measure (PRF-A; Table 7). Calculating the estimated differences between groups based on the model results, we found that the TI group gained an estimated 3.72 standard score points on word reading (2.31 + 0.07*(+1) + 1.34*(+1)) whereas the PI group's estimated gain is 1.04 standard score points (2.31 + 0.07*(+1) + 1.34*(–1)), $t(199) = 2.01$, $p < .05$, $d = .43$. Similarly, students who were tutored by teachers gained an estimated 38.42 wcpm on PRF-A (the alternate passage reading fluency measure) compared with their paraeducator-tutored peers, who were estimated to have gained 30.08 wcpm, $t(199) = 2.04$, $p < .05$, $d = .43$.

Research Questions 3 and 4: Unique Effects of Treatment-Related Variables

Table 8 displays zero-order correlations among classroom instruction variables; student pretest–posttest gains; and, for treatment students, treatment-related variables (correlations are unadjusted for classroom or dyad variance; the upper diagonal reports control students' intercorrelations whereas the lower diagonal displays treatment students' intercorrelations.) To review, the purpose of our third research question was to determine the unique effects of tutoring fidelity (treatment "dosage") and word-level decoding complexity (mean DCI values of each dyad's unique words read in *Quick Reads* passages) on treatment students' pretest–posttest gains. Decoding complexity was of interest specifically because the influence of word-level complexity in repeated reading interventions remains unclear (Chard, Vaughn, & Tyler, 2002; Kuhn & Stahl, 2003). Further, we did not include other text characteristics in our models primarily because we found mean DCI values to be extremely highly correlated with the SFI and total number of unique words read (see Table 8), indicating that variability in DCI values for these texts is redundant with variability in SFI values and word uniqueness. Finally, we were interested in whether these two treatment-related variables (tutoring fidelity and text characteristics) contributed uniquely to individual differences in gains once core pretest skills were controlled for, including receptive language, rapid naming, and word reading, variables that account for fluency and comprehension performance in the early grades (Adolf et al., 2006; Cutting & Scarborough, 2006). In other words, we wished to know whether tutoring fidelity and text characteristics could explain differences in response to intervention above and beyond baseline skill levels. (Recall that there were no significant differences between groups on tutoring fidelity and thus treatment implementer group type is not included in these models.) Thus, we specified multilevel models in which five predictors (the two treatment-related predictors and three pretest skills) were used to predict treatment students' pretest–posttest gains.The results of our models are shown in Table 9. Results revealed that tutoring fidelity uniquely predicts gains in word reading and passage

TABLE 8
Intercorrelations of Classroom Instruction, Pretest–Posttest Gains, and Treatment-Related Variables

Measure	*1.*	*2.*	*3.*	*4.*	*5.*	*6.*	*7.*	*8.*	*9.*	*10.*	*11.*	*12.*	*13.*	*14.*	*15.*	*16.*	*17.*	*18.*
Classroom Instruction (Daily Minutes)																		
1. Word Reading	—	–.16	**.23**	**–.34**	.09	.06	.12	.01	–.14	–.06	.04	.05	.04					
2. Text Reading (Oral, Silent)	**–.32**	—	.14	**.20**	–.17	.07	–.08	.06	–.12	–.07	–.05	.17	.04					
3. Vocabulary, Comprehension	.16	.03	—	.16	.00	–.15	.07	.00	–.06	.01	–.05	–.02	–.01					
4. Writing	**–.23**	**.24**	**.23**	—	–.02	.00	–.03	–.15	–.03	–.07	–.16	–.13	–.09					
Word-Level Skills (Pretest–Posttest Gains)																		
5. Single Letter Sounds	–.06	–.02	.10	–.01	—	**.22**	.04	.08	.09	.10	**.29**	–.09	.04					
6. Letter Combo Sounds	.10	.16	–.09	**.20**	**.26**	—	**.21**	.19	–.10	–.02	**.29**	.02	.06					
7. Word Reading	**.27**	–.04	.05	.01	–.03	**.20**	—	**.21**	.18	**.24**	**.21**	**.21**	**.32**					
Reading Fluency (Pretest–Posttest Gains)																		
8. PRF-U	–.02	.09	–.10	–.04	.05	**.30**	**.29**	—	**.45**	**.58**	**.31**	.03	.25					
9. PRF-A	.03	–.01	.00	–.04	**.24**	.19	**.31**	**.63**	—	**.57**	.16	–.08	.10					
10. Fluency Rate	–.01	.06	–.12	–.04	.05	.17	**.20**	**.68**	**.69**	—	.13	–.14	**.27**					

(continued)

TABLE 8 *(Continued)*

Measure	*1.*	*2.*	*3.*	*4.*	*5.*	*6.*	*7.*	*8.*	*9.*	*10.*	*11.*	*12.*	*13.*	*14.*	*15.*	*16.*	*17.*	*18*
Comprehension (Pretest–Posttest Gains)																		
11. Cloze-Task	.09	.01	.06	.06	.11	**.30**	**.55**	**.52**	**.37**	**.40**	—	–.02	.04					
12. Mult Choice	.02	.00	.07	–.09	.13	.00	–.17	.03	–.10	.06	–.02	—	**.60**					
13. Mult Choice, FB	.06	–.07	–.03	–.10	.13	.17	.14	**.23**	**.21**	**.40**	**.24**	**.58**	—					
Treatment-Related Variables																		
14. Tutoring Fidelity	.04	.12	.08	.04	–.12	–.07	.16	.17	**.21**	.14	–.01	.02	.06	—				
15. Flesch-Kincaid Grade Level	–.18	–.19	.00	.15	.06	**–.23**	–.07	**–.40**	.11	–.06	**–.22**	–.15	–.08	–.08	—			
16. Total Unique Words	–.09	–.09	–.07	.10	.01	**–.26**	–.07	**–.29**	.11	–.04	–.17	**–.23**	–.13	–.06	**.82**	—		
17. SFI, Unique Words	.11	.11	.10	–.10	.02	**.26**	.08	**.25**	–.11	.02	.16	**.23**	.12	.03	**–.76**	**–.99**	—	
18. DCI, Unique Words	–.15	.15	–.04	.15	.03	**–.27**	–.08	**–.37**	.11	–.06	**–.22**	–.18	–.12	–.07	**.96**	**.93**	**–.89**	—

Note. Treatment group ($n = 98$) correlations in lower diagonal ($n = 96$ for classroom variables due to missing teacher survey); Control group ($n = 104$) correlations in upper diagonal ($n = 102$ for classroom variables due to missing teacher survey). Single Letter Sounds = number of correctly read sounds of isolated letters (25 possible); Letter Combo Sounds = number of correctly read sounds of letter combinations (42 possible); Word Reading = standard score of Word Identification subtest from Woodcock Reading Mastery Test-Revised/Normative Update; PRF = passage reading fluency as measured by uniform (U) and alternate (A) passages drawn from the Dynamic Indicators of Basic Early Literacy Skills ORF benchmarks; Fluency Rate = standard score of Rate subtest from Gray Oral Reading Test Version 4 (GORT-4); Cloze-Task = standard score of WRMT-R/NU Passage Comprehension subtest; Mult Choice = standard score of Comprehension subtest from GORT-4; Mult Choice, FB = adapted, fluency-based standard score of Comprehension subtest from GORT-4; Tutoring Fidelity = 5-point rating of dyad's tutoring fidelity (higher scores = better performance); Flesch-Kincaid Grade Level = mean passage grade level as rated by Microsoft Word 2003 on the first passage of each book in *Quick Reads* level covered; Total Unique Words = total non-repeating words read in *Quick Reads* passages; SFI = standard frequency index of dyad's unique words read in *Quick Reads* passages (higher scores = more frequent words); DCI = 11-point decoding complexity rating of dyad's unique words read in *Quick Reads* passages (higher scores = more difficult words). Pearson's r used. Correlations in bold face are significant at the .05 level.

TABLE 9
Unique Effects of Treatment-Related Variables on Pretest–Posttest Gains

	Fixed Effects																		Random Effects			
	Intercept (Gain)			Pretest PPVT			Pretest RAN			Pretest WR			Tutoring Fidelity			DCI, Unique Words			Dyads			Residueal
Measure	*M*	*SE*	*t*(46)	*M*	*SE*	*t*(92)	*M*	*SE*	*t*(92)	*M*	*SE*	*t*(92)	*M*	*SE*	*t*(46)	*M*	*SE*	*t*(46)	*Var*	χ^2 (46)	%	*Var*
Word-Level Skills																						
Single Letter Sounds	0.93	0.21	4.46***	0.00	0.01	–0.04	0.02	0.02	1.00	–0.02	0.03	–0.48	–0.69	0.51	–1.37	0.21	0.89	0.24	0.34	58.35	.09	3.57
Letter Combo Sounds	8.50	0.66	12.86***	0.05	0.04	1.25	0.09	0.07	1.32	–0.26	0.07	–3.84***	–1.39	1.45	–0.95	–7.63	2.02	–3.78***	8.03	78.42**	.23	26.77
Word Reading	2.49	0.49	5.05***	0.05	0.04	1.28	–0.03	0.04	–0.87	–0.36	0.10	–3.52***	2.77	1.35	2.06*	–1.36	2.39	–0.57	0.02	37.98	.00	30.69
Reading Fluency																						
PRF–U	46.45	1.78	26.05***	0.26	0.10	2.65**	0.52	0.20	2.63**	0.27	0.31	–0.87	11.31	4.07	2.78**	–34.26	6.54	–5.24***	0.78	44.75	.00	
PRF–A	34.52	2.10	16.42***	–0.06	0.13	–0.47	0.42	0.21	1.95	0.10	0.36	0.26	13.87	4.96	2.80**	12.31	7.28	1.69	21.53	54.51	.05	389.27
Fluency Rate	13.16	1.35	11.60***	0.06	0.07	0.86	0.24	0.12	1.93	0.01	0.17	0.08	5.36	3.04	1.76	–2.37	3.97	–0.60	4.83	53.24	.04	116.13
Comprehension																						
Cloze– Task	2.99	0.55	5.41***	0.09	0.05	1.83	0.02	0.05	0.37	–0.17	0.13	–1.31	–0.11	1.81	–0.06	–6.27	1.99	–3.14**	0.03	37.10	.00	39.58
Mult Choice	7.60	1.31	5.80***	0.09	0.07	1.29	–0.06	0.13	–0.44	0.24	0.18	1.31	1.01	3.29	0.31	–11.61	6.20	–1.87	0.72	49.17	.00	168.01
Mult Choice, FB	9.03	0.89	10.15***	0.07	0.06	1.21	0.07	0.09	0.76	0.03	0.13	0.21	1.99	2.86	0.70	–4.82	3.49	–1.38	0.24	45.97	.00	82.71

Note. Treatment students' ($n = 98$) pretest–posttest gains analyzed. Predictors were grand-mean centered and include: Pretest PPVT = pretest standard score of Peabody Picture Vocabulary Test Version 3; Pretest RAN = pretest standard score of Letter Naming subtest from Rapid Automatized Naming and Stimulus Subtests; Pretest WR = pretest standard score of Word Identification subtest from Woodcock Reading Mastery Test-Revised/Normative Update (WRMT-R/NU); Tutoring Fidelity = 5-point mean rating of each dyad's tutoring fidelity (higher scores = better performance); DCI = 11-point decoding complexity mean rating of dyad's unique words read in *Quick Reads* passages (higher scores = more difficult words). Gains analyzed include: Single Letter Sounds = number of correctly read sounds of isolated letters (25 possible); Letter Combo Sounds = number of correctly read sounds of letter combinations (42 possible); Word Reading = standard score of Word Identification subtest from WRMT-R/NU; PRF = passage reading fluency as measured by uniform (U) and alternate (A) passages drawn from the Dynamic Indicators of Basic Early Literacy Skills ORF benchmarks; Fluency Rate = standard score of Rate subtest from Gray Oral Reading Test Version 4 (GORT-4); Cloze-Task = standard score of WRMT-R/NU Passage Comprehension subtest; Mult Choice = standard score of Comprehension subtest from GORT-4; Mult Choice, Fluency-Based = adapted, fluency-based standard score of Comprehension subtest from GORT-4

* $p < .05$. ** $p < .01$. *** $p < .001$.

reading fluency (both PRF-U and PRF-A), whereas decoding complexity uniquely predicts gains in one word-level skill (combination letter sounds), one measure of reading fluency (PRF-U), and one comprehension outcome (cloze task). Holding all other variables constant at their mean, students from dyads with one standard deviation higher than average tutoring fidelity were predicted to have a 2.77-point advantage on word reading gains, an 11.31-wcpm advantage on PRF-U gains, and a 13.87-wcpm advantage on PRF-A gains. Interpreting the decoding complexity effects (recall that higher scores indicate more complex words) reveals that students from dyads who read texts that were one standard deviation more complex than average were predicted to have lower gains in combination letter sounds (7.63 sounds lower), uniform passage reading fluency (PRF-U; 34.26 wcpm lower), and the comprehension cloze task measure (6.27 standard score points lower), holding all other variables constant.

DISCUSSION

The present study examined the efficacy of a supplemental, repeated reading intervention with added word-level instruction for second and third graders with lower fluency skills, with an added comparison of teacher and paraeducator implementers. Results showed that, on average, students benefited from *Quick Reads* instruction in word level, fluency, and comprehension gains, findings consistent with those reported previously for an earlier second- and third-grade fluency intervention (Vadasy & Sanders, 2008b). Our model results for the present study show specifically that tutored students had significantly higher pretest–posttest gains in one word-level measure (correctly identifying two- and three-letter spelling patterns), reading fluency (the uniform passage reading fluency measure and the norm-referenced GORT fluency rate measure), and passage reading comprehension (the cloze task measure and the fluency-based scoring of the GORT multiple choice questions).

Unlike the large word-level effects achieved by Wolf et al. (2000) for a similar fluency and word- level intervention (also provided to second and third graders in the early stage of fluency development), the word reading effects we observed in our earlier study (Vadasy & Sanders, 2008b) and in the present study were low to modest. In the present study, gains in knowledge of taught letter patterns did not appear to transfer to word reading outcomes. One explanation for the lack of treatment effects for word reading in this study consistent with the *simple reading fluency model* (Schwanenflugel et al., 2006) is that intensity or quality of instruction allocated to word reading accuracy in the treatment did not allow students to develop sufficient word reading fluency to advance beyond the “decoding reader stage” (Wolf, 2007). Treatment effects in fluency and comprehension were small to moderate and were possibly constrained by inadequate word reading accuracy and efficiency: At pretest, both treatment and control students averaged in approximately the 10th and 30th percentiles in reading fluency and cloze-task passage reading comprehension. At posttest, treatment students averaged in the 30th and 40th percentiles in fluency and cloze task comprehension, whereas controls averaged in the 25th and 35th percentiles, respectively. The vocabulary, background knowledge, and language proficiency of students in this study may have also limited growth in comprehension. Treatment students averaged in the lower 30th percentile in receptive language at pretest, and 30% were English Language Learners. The significant contribution of decoding complexity of words in the *Quick Reads* passages to treatment students' gains in fluency and comprehension also suggests that inefficient decoding may have impeded growth in these higher level reading skills.

Passage Reading Fluency Measurement

Recall that students read two DIBELS benchmark grade-level text passages for which passage reading fluency rates were computed: one uniform (PRF-U) and one alternate (PRF-A) passage. Treatment students made significantly greater gains than controls when measured on the uniform passage, but significant differences were not detected on the alternate passage gain. We speculate that this disparity is due to lack of power (the treatment effect sizes are small across all fluency measures), the likely nonequivalence of the alternate form passages used, and possible passage memory effects for the uniform passage used. Across conditions and grades, student performance on the uniform and alternate passages differed significantly at pretest by only 3.49 wcpm: $t(201) = 4.63$, $p < .001$ ($M = 43.71$ and $M = 40.22$ for PRF-U and PRF-A, respectively). Comparatively, the difference widened to 13.74 wcpm by posttest, $t(201) = 14.50$, $p < .001$ (PRF-U $M = 72.48$ and PRF-A $M = 86.21$). This finding is in alignment with Jenkins et al. (2005), who found larger gains on uniform passages at 5-week retest compared to nonuniform passages, indicating some passage memory effects were present in our sample (recall that students were also exposed to the uniform passage at midtest, which was approximately 5 weeks after pretest, although midtest data were not analyzed in the present study). In addition, it is likely that alternate form passage gains are prone to measurement error from passage nonequivalence effects. These issues of memory and nonequivalence effects are qualified by the correlations between the uniform and alternate measures at each test wave, which were highly similar: $r = .83$ and $r = .84$ at pretest and posttest, respectively. This would suggest that the rank ordering of children was consistent at each test wave irrespective of passage type (i.e., students who are high on one passage are likely to be high on the other passage). Although treatment students made significantly greater gains on the standard GORT fluency rate subtest, both fluency treatment effects were small and similar in magnitude to those reported previously (Vadasy & Sanders, 2008b).

Comprehension Measurement

The treatment effects we detected were small for both cloze-task (WRMT–R/NU Passage Comprehension subtest) and multiple-choice comprehension formats (adapted scoring for GORT-4 Comprehension subtest). It is noteworthy that significant treatment effects were detected only for the adapted, fluency-based administration of the multiple-choice comprehension measure (scores based only on those GORT-4 passages students read with some fluency) but not for the standard administration of that subtest (which requires the child to achieve a separate basal and ceiling for fluency resulting in test questions that may be passage independent). This finding can be considered in light of concerns raised by Keenan and Betjemann (2006) regarding validity of standard administration of the GORT Comprehension subtest, which allows students to answer multiple-choice items without having actually read the corresponding text passage. In our findings, the treatment impact on the cloze-task gains (d = .33) was most similar to the fluency-based (adapted) multiple-choice gains (d = .31), even though the correlation between each measure was only moderate: across groups, correlations between the cloze task and fluency-based scoring of the multiple-choice task are r = .59 and .61 at pretest and posttest, respectively. In contrast, the correlations are r = .45 and .41 between the cloze task and the standard scoring of the multiple choice task at pretest and posttest, respectively.

On the other hand, if one assumes that standard scoring (rather than the fluency-based scoring) of the multiple-choice task is a more valid measure, the disparity in comprehension treatment effects (d = .33 for cloze task and d = –.02 for multiple choice) aligns with the simple view of early stages of reading development, and the influence of testing formats on comprehension assessment (Cutting & Scarborough, 2006; Nation & Snowling, 1997; Spear-Swerling, 2004). As shown in Table 8, we found that treatment students' gains in the standard scoring of the multiple choice task were not correlated with any other gain except the fluency-based scoring of the same task, whereas gains in the fluency-based scoring of the multiple choice task were correlated with all three reading fluency (and the cloze-task comprehension) measures. Moreover, gains measured by the cloze task were correlated with both word-level and reading fluency gains, suggesting the cloze-task format is heavily influenced by word reading skills (Keenan, Betjemann, & Olson, 2008). In sum, we share concerns raised by others (Cutting & Scaroborough, 2006; Keenan & Betjemann, 2006; Keenan et al., 2008; Nation & Snowling, 1997) regarding the relationship of comprehension assessment task formats and the skills thought to be measured in the early grades and stages of reading development.

Treatment-Related Outcomes

A question we have addressed in previous studies is the most effective utilization of paraeducators to supplement early reading skills in high needs schools. In this study, treatment students who were tutored by certificated teachers gained more on word reading and alternate passage reading fluency than students tutored by paraeducators. Teachers may have had an advantage in word reading instruction because this skill requires more skillful scaffolding and individualization, unlike the more straightforward procedures for the fluency practice. Teachers likely brought more skills and experience to the word-level layer of the intervention than did the paraeducators. However, schools' variation in pull-out patterns between teacher and paraeducator groups may also account for this difference. Students tutored by paraeducators were more often pulled out during their classroom reading block when they might have received more skilled and added word reading instruction. The single difference between implementer types in word reading (with the fluency difference between implementer type no longer significant when controlling for fidelity), together with the average teaching experience of the teacher tutors (averaging 14 years), suggests that paraeducators were able to effectively deliver this type of fluency instruction and practice. Paraeducators may have been better able to teach word-level skills had they been provided with additional training or more coaching to better match the intensity of this instruction to individual student needs.

Our treatment-related multilevel regression models suggest that both fidelity of tutoring implementation (i.e., treatment "dosage") and intervention text features (i.e., decoding complexity of words in the texts read) uniquely accounted for differences in reading fluency gains. Controlling for treatment fidelity and pretest receptive language, rapid naming, and word reading skills, we found that students who read more orthographically complex words (i.e., less regular, more two- to three-letter spelling patterns, more syllables) had, on average, lower alphabetic, fluency, and comprehension gains. This finding is difficult to explain, although it must be considered in light of the overall high level of decodability of words in the passages students read. Our text analyses indicated that 77% of all unique words in the texts could be considered "decodable." The texts read by the majority of students in the study provided exposure to a broad range of decodable word types. In this continuum, a substantial proportion (21–34%) of more challenging, multisyllabic words were repeated in different contexts. Further, across all *Quick Reads* passages, each word occurred approximately eight times, on average. According to Zeno et al. (1995), the *Quick Reads* texts reflect an unusually high standard frequency value, and therefore this intervention exposed students to a large proportion of high-frequency words, in repeated occurrences across passages. This text feature may result in a high level of transfer to other grade-level texts, as is seen in our study's fluency gains. It may be that tutors did not adequately scaffold word reading as text difficulty increased, and student word reading skills may not have developed adequately to keep pace with the gradually more challenging words they encountered, limiting the gains for fluency and comprehension. That is, greater tutor sensitivity to the word features of the passages and appropriate scaffolding may have allowed students' word reading skills to keep pace with the already considerate and gradual increase in word complexity in these texts.

Limitations

This study had several limitations. First, although the intervention used in this study may be primarily characterized as repeated reading, a portion of each tutoring session included word-level instruction and scaffolding. Although many repeated reading interventions described in the literature do not include this feature, it is one of the extension options in the *Quick Reads* teacher manual. Second, information on classroom teachers' reading instruction from our teacher survey may be highly inaccurate due to errors in self-report (see Burstein et al., 1995) and thus should be treated tentatively (it is also noteworthy that, for controls, teacher instruction variables did not correlate with any pretest–posttest gains, as shown in Table 8).

Third, students entered this study with a wide range of pretest fluency levels that reflected teacher referral patterns (i.e., not all students were below average in fluency). Nevertheless, students ranged from 10th to 60th percentile on passage reading fluency performance, similar to students typically served with repeated reading intervention (Faulkner & Levy, 1999; Hasbrouck, Ihnot, & Rogers, 1999). Fluency was measured at pretest and posttest only, and repeated measures of oral reading fluency would have allowed for analysis of slope.

Fourth, students in this study were randomly assigned to dyads within grade and school (across classrooms), whereas in typical school applications, students may be systematically paired with a peer of similar skill level within a classroom. As such, the study design does not permit us to inform how homogeneity of dyad skill level may influence student outcomes. Nevertheless, our findings show that discrepancies in students' skill levels within dyads were minimal (as shown by small and mostly nonsignificant intraclass correlations for dyads on pretests and pretest–posttest gains in Tables 6, 7, and 9), and therefore our dyads appear mostly homogeneous despite the randomized pairings we employed.

A fifth limitation is that treatment dyads were systematically placed into one of two different text levels based on their grade level. As such, the range of text decoding complexity (and therefore the relationships found between decoding complexity and fluency gains) is restricted to the range of texts students actually read. In other words, we cannot draw inferences about the relationship between intervention decoding complexity and fluency gains for the lower and upper extremes of the 11-point decoding complexity rating scale we employed.

A final limitation of this study is that a matched-groups design was used to compare implementer type (certificated teachers vs. paraeducators). Specifically, implementer group assignment is confounded with school site, and as such differences between teacher and paraeducator implementer groups may reflect differences between two sets of schools. Nevertheless, groups were rigorously matched on academic risk and school demographic variables to increase confidence in the findings for this policy-relevant question.

CONCLUSION

In summary, the findings from this study suggest that the core *Quick Reads* repeated reading instruction achieved significant improvements in fluency and comprehension. Improvements for the low-skilled students in this study were modest, however, and as others have cautioned, more intensive or extended intervention appears needed to accomplish grade-level fluency performance. Findings from the treatment comparison suggest that paraeducators may be as effective as certificated teachers in boosting students' reading skills, in particular the fluency and comprehension skills that are the typical focus of repeated reading intervention. This finding may increase confidence in using these school staff to provide low-skilled students with opportunities for additional reading practice. Schools that assign paraeducator staff to supplement fluency instruction for students with low reading rates might also enable skilled classroom teachers to effectively differentiate classroom comprehension and word reading instruction for these students.

Finally, our results showed that teachers were more effective than paraeducators in teaching word-level skills. Although the word reading layer of instruction in this study represented only a brief portion of the tutoring sessions—and was quite individualized and varied among students—this finding suggests the need for skilled teacher instruction to differentiate remediation in word reading skills. As students reach second and third grade, word reading demands in texts increase and require students to possess a wider range of word-level skills rather than simple application of basic decoding strategies. Indeed, successful word reading in these grades is characterized by increasing automaticity, sensitivity to regular and consistent spelling patterns, rapid recognition of high-frequency sight words, and an emerging strategy for reading multisyllable words. The complexity of increasingly multilayered word reading skills suggests that its instruction may be best provided by skilled teachers, particularly when embedded in a fluency intervention.

ACKNOWLEDGMENTS

Grant R305G040103A from the Institute of Education Sciences, U.S. Department of Education, supported this research. We especially acknowledge Sueanne Sluis for coordinating the intervention and the classroom observations. We also thank Kathryn Compton for coordinating student assessments, and we thank testers, observers, and data entry staff for their work on the project, including Katy Compton, Alison Fiorito, Christine Haas, Rayma Haas, Gwenneth Kunde, Becky Logan, Ruthanne McPhaden, Devon Parris, Laura Root, and Ashley Wagenfuhr. Last but not least, we are most grateful to the children, tutors, teachers, and administrative staff of the schools for their support and participation in this study.

REFERENCES

Abrami, P. C., Lou, Y., Chambers, B., Poulsen, C., & Spence, J. C. (2000). Why should we group students within-class for learning? *Educational Research and Evaluation, 6*, 158–179.

Adolf, S. M., Catts, H. W., & Little, T.D. (2006). Should the simple view of reading include a fluency component? *Reading and Writing, 19*, 933–958.

Allington, R. L. (1977). If they don't read much, how they ever gonna get good? *Journal of Reading, 21*, 57–61.

Anderson, R. C., Hiebert, E. H., Scott, J. A., & Wilkinson, I. A. (1985). *Becoming a nation of readers: The report of the Commission on Reading*. Washington, DC: National Academy of Education.

Bowey, J. (1986). Syntactic awareness in relation to reading skill and ongoing reading comprehension monitoring. *Journal of Experimental Child Psychology, 41*, 282–299.

Burstein, L., McDonnell, L. M., Van Winkle, J. T., Ormseth, H., Mirocha, J., & Guiton, G. (1995). *Validating national curriculum indicators*. Santa Monica, CA: RAND.

Carlisle, J. F., & Rice, M. S. (2004). Assessment of reading comprehension. In C. A. Stone, E. R. Silliman, B. J. Ehren, & K. Apel (Eds.), *Handbook of language and literacy: Development and disorders* (pp. 521–540). New York: Guilford.

Catts, H. W., Fey, M. E., Zhang, Z., & Tomblin, J. B. (1999). Language basis of reading and reading disabilities: Evidence from a longitudinal investigation. *Scientific Studies of Reading, 3*, 331–361.

Chard, D. J., Vaughn, S., & Tyler, B. (2002). A synthesis of research on effective interventions for building reading fluency with elementary students with learning disabilities. *Journal of Learning Disabilities, 35*, 386–406.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Compton, D. L., Appleton, A. C., & Hosp, M. K. (2004). Exploring the relationship between text-leveling systems and reading accuracy and fluency in second-grade students who are average and poor decoders. *Learning Disabilities Research & Practice, 19*, 176–184.

Cutting, L. E., & Scarborough, H. S. (2006). Prediction of reading comprehension: Relative contributions of word recognition, language proficiency, and other cognitive skills can depend on how comprehension is measured. *Scientific Studies of Reading, 10*, 277–299.

Dahl, P. R. (1979). An experimental program for teaching high speed word recognition and comprehension skills. In J. E. Button, T. C. Lovitt, & T. D. Rowlands (Eds.), *Communication research in learning disabilities and mental retardation* (pp. 33–65). Baltimore: University Park Press.

Dunn, L. M., & Dunn, L. (1997). *Peabody Picture Vocabulary Test–Revised (PPVT–R)*. Circle Pines, MN: American Guidance Service.

Ehrenberg, R., & Brewer, D. (1994). Do school and teacher characteristics matter? Evidence from high school and beyond. *Economics of Education Review, 14,* 1–23.

Elbaum, B., Vaughn, S., Hughes, M., & Moody, S. (2000). 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 Psychology, 92*, 605–619.

Farr, R., Pritchard, R., & Smitten, B. (1990). A description of what happens when an examinee takes a multiple-choice reading comprehension test. *Journal of Educational Measurement, 27*, 209–116.

Faulkner, H. J., & Levy, B. A. (1999). How text difficulty and reader skill interact to produce differential reliance on word and context overlap in reading transfer. *Journal of Experimental Child Psychology, 58,* 1–24.

Ferguson, R., & Ladd, H. (1996). How and why money matters: An analysis of Alabama schools. In H. Ladd (Ed.), *Holding schools accountable* (pp. 265–298). Washington, DC: Brookings Institution.

Fletcher, J. M. (2006). Measuring reading comprehension. *Scientific Studies of Reading, 10*, 323–330.

Francis, D. J., Fletcher, J. M., Catts, H. W., & Tomblin, J. B. (2005). Dimensions affecting the assessment of reading comprehension. In S. G. Paris & S. A. Stahl (Eds.), *Children's reading comprehension and assessment* (pp. 369–394). Mahwah, NJ: Erlbaum.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Karns, K. (1998). High-achieving students' interactions and performance on complex mathematical tasks as a function of homogeneous and heterogeneous pairings. *American Educational Research Journal, 35*, 227–268.

Fuchs, L. S., Fuchs, D., & Maxwell, L. (1988). The validity of informal measures of reading comprehension. *Remedial and Special Education, 9*, 20–28.

Gerber, S. B., Finn, J. D., Achilles, C. M., & Boyd-Zaharias, J. (2001). Teacher aides and students' academic achievement. *Educational Evaluation and Policy Analysis, 23*, 123–143.

Good, R. H., & Kaminski, R. A. (2002). *Dynamic Indicators of Basic Early Literacy Skills* (6th ed). Eugene, OR: Institute for the Development of Educational Achievement.

Gough, P. B., Hoover, W., & Peterson, C. L. (1996). Some observations on the simple view of reading. In C. Cornoldi & J. Oakhill (Eds.), *Reading comprehension difficulties* (pp. 1–13). Mahwah, NJ: Erlbaum.

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

Greenwald, R., Hedges, L., & Lane, R. (1996). The effect of school resources on student achievement. *Review of Educational Research, 66*, 361–396.

Grek, M. L., Mathes, P. G., & Torgesen, J. K. (2003). Similarities and differences between experienced teachers and trained paraprofessionals. In S. Vaughn & K. L. Briggs (Eds.), *Reading in the classroom: Systems for the observation of teaching and learning* (pp. 267–296). Baltimore: Brookes.

Hasbrouck, J. E., Ihnot, C., & Rogers, G. H. (1999). "Read Naturally": A strategy to increase oral reading fluency. *Reading Research and Instruction, 39,* 27–38.

Hasbrouck, J., & Tindal, G. A. (2006). ORF norms: A valuable assessment tool for reading teachers. *The Reading Teacher, 59*, 636–644.

Hiebert, E. H. (2003). *Quick reads*. Parsippany, NJ: Pearson Learning.

Hiebert, E. H. (2005). The effects of text difficulty on second graders' fluency development. *Reading Psychology, 26*, 183–209.

Hiebert, E. H. (2008). The (mis)match between texts and students who depend on schools to become literate. In E. H. Hiebert & M. Sailors (Eds.), *Finding the right texts: What works for beginning and struggling readers* (pp. 1–20). New York: Guilford.

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing, 2*, 127–160.

Jenkins, J. R., Fuchs, L. S., van den Broek, P., Espin, C., & Deno, S. L. (2003). Sources of individual differences in reading comprehension and reading fluency. *Journal of Educational Psychology, 95,* 719–729.

Jenkins, J. R., Zumeta, R., Dupree, O., & Johnson, K. (2005). Measuring gains in reading ability with passage reading fluency. *Learning Disabilities Research & Practice, 20*, 245–253.

Joshi, R., & Aaron, P.G. (2000). The component model of reading: Simple view of reading made a little more complex. *Reading Psychology, 21*, 85–97.

Kain, J., & Singleton, K. (1996). Equality of education revisited. *New England Economic Review,* May–June, 87–111.

Kame'enui, E. J., Simmons, D. C., Good, R. H., & Harn, B. A. (2001). The use of fluency-based measures of early identification and evaluation of intervention efficacy in schools. In M. Wolf (Ed.), *Dyslexia, fluency, and the brain* (pp. 307–332). Timonium, MD: York Press.

Katz, S., Lautenschlager, G., Blackburn, A. B., & Harris, F. (1990). Answering reading comprehension items without passages on the SAT. *Psychological Science, 1*, 122–127.

Keenan, J. M., & Betjemann, R. S. (2006). Comprehending the Gray Oral Reading Test without reading it: Why comprehension tests should not include passage-independent items. *Scientific Studies of Reading, 10*, 363–380.

Keenan, J. M., Betjemann, R. S., & Olson, R. K. (2008). Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific Studies of Reading, 12,* 281–300.

Kuhn, M. R., & Stahl, S. A. (2003). Fluency: A review of developmental and remedial practices. *Journal of Educational Psychology, 95*, 3–21.

Lou, Y., Abrami, P. C., & Spence, J. C. (2000). Effects of within-class grouping on student achievement: An exploratory model. *The Journal of Educational Research, 94*, 101–112.

Lou, Y., Abrami, P. C., Spence, J. C., Poulsen, C., Chambers, B., & d'Apollonia, S. (1996). Within-class grouping: A meta-analysis. *Review of Educational Research, 66,* 423–458.

Menon, S., & Hiebert, E. H. (1999). *Literature anthologies: The task for first grade readers*. (Report No. CIERA–R–1–009). Ann Arbor, MI: Center for the Improvement of Early Reading Achievement. (ERIC Documentation Service No. ED436754)

Nation, K., & Snowling, M. (1997). Assessing reading difficulties: The validity and utility of current measure of reading skill. *British Journal of Educational Psychology, 67*, 359–370.

Nelson, S. R. (1999). *Designing state and local policies for the professional development of instructional paraeducators.* Portland, OR: Northwest Regional Educational Laboratory.

Perfetti, C. (1977). Language comprehension and fast decoding: Some psycholinguistic prerequisites for skilled reading comprehension. In J. T. Guthrie (Ed.), *Cognition, curriculum, and comprehension* (pp. 20–41). Newark, DE: International Reading Association.

Perfetti, C. A. (1985). *Reading ability.* New York: Oxford University Press.

Phelps, E., & Damon, W. (1989). Problem solving with equals: Peer collaboration in a context for learning mathematics and spatial concepts. *Journal of Educational Psychology, 81*, 639–646.

Pierce, M. E., Katzir, T., Wolf, M., & Noam, G. (2007). Clusters of second and third grade dysfluent urban readers. *Reading and Writing: An Interdisciplinary Journal, 20,* 885–907.

Pikulski, J. (1994). Preventing reading failure: A review of five effective programs. *The Reading Teacher, 48*, 30–39.

Pryczak, F. (1972). Objective evaluation of the quality of multiple-choice items designed to measure comprehension of reading passages. *Reading Research Quarterly, 8,* 62–72.

Rashotte, C. A., MacPhee, K., & Torgesen, J. K. (2001). The effectiveness of a group reading instruction programs with poor readers in multiple grades. *Learning Disability Quarterly, 24*, 119–134.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., & Congdon, R. T. (2004). *HLM 6: Hierarchical linear and nonlinear modeling*. Lincolnwood, IL: Scientific Software International.

Rosenthal, R., & Rosnow, R. L. (1984). *Essentials of behavioral research: Methods and data analysis*. New York: McGraw-Hill.

Rowan, B. (2002). What large-scale survey research tells us about teacher effects on student achievement: Insights from the Prospects Study of Elementary Schools. *Teachers College Record, 104*, 1525–1567.

Rupley, W. H., Willson, V. L., & Nichols, W. D. (1998). Exploration of the developmental components contributing to elementary school children's reading comprehension. *Scientific Studies of Reading, 2*, 143–158.

Samuels, S. J. (1979). The method of repeated readings. *The Reading Teacher, 32,* 403–408.

Savage, R. (2001). The 'simple view' of reading: Some evidence and possible implications. *Educational Psychology in Practice, 17*, 17–22.

Schwanenflugel, P. J., Meisinger, E. B., Wisenbaker, J. M., Kuhn, M. R., Strauss, G. P., & Morris, R. D. (2006). Becoming a fluent and automatic reader in the early elementary years. *Reading Research Quarterly, 41*, 496–522.

Seigneuric, A., & Ehrlich, M. (2005). Contribution of working memory capacity to children's reading comprehension: A longitudinal investigation. *Reading and Writing, 18*, 617–656.

Spear-Swerling, L. (2004). Fourth graders' performance on a state-mandated assessment involving two different measures of reading comprehension. *Reading Psychology, 25*, 121–148.

Stanovich, K. E. (1980). Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly, 16*, 32–71.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21*, 360–407.

Torgesen, J. K., Rashotte, C. A., & Alexander, A. W. (2001). Principles of fluency instruction in reading: Relationships with established empirical outcomes. In M. Wolf (Ed.), *Dyslexia, fluency, and the brain* (pp. 333–356). Timonium, MD: York Press.

Vadasy, P. F., & Sanders, E. A. (2008a). Benefits of repeated reading intervention for low-achieving fourth- and fifth-grade students. *Remedial and Special Education, 29*(4), 235–249.

Vadasy, P. F., & Sanders, E. A. (2008b). Repeated reading intervention: Outcomes and interactions with readers' skills and classroom instruction. *Journal of Educational Psychology, 100*(2), 272–290.

Wasik, B. A., & Slavin, R. E. (1993). Preventing early reading failure with one-to-one tutoring: A review of five programs. *Reading Research Quarterly, 28*, 178–200.

Wiederholt, J. L., & Bryant, B. R. (2001). *Gray Oral Reading Tests–4*. Austin, TX: Pro-Ed.

Wolf, M. (2007). *Proust and the squid: The story and science of the reading brain.* New York: HarperCollins.

Wolf, M., & Denckla, M. B. (2005). *RAN/RAS: Rapid automatized naming and rapid alternating stimulus tests.* Austin, TX: Pro-Ed.

Wolf, M., & Katzir-Cohen, T. (2001). Reading fluency and its intervention. *Scientific Studies of Reading, 5*, 211–239.

Wolf, M., Miller, L., & Donnelly, K. (2000). Retrieval, automaticity, vocabulary elaboration, orthography (RAVE-O): A comprehensive, fluency-based reading intervention program. *Journal of Learning Disabilities, 33*, 375–386.

Woodcock, R. (1998). *Woodcock Reading Mastery Test–Revised/Normative Update*. Circle Pines, MN: American Guidance Service. (Original work published 1987)

Zeno, S. M., Ivens, S. H., Millard, R. T., & Duvvuri, R. (1995). *The Educator's Word Frequency Guide*. Touchtone Applied Science Associates.