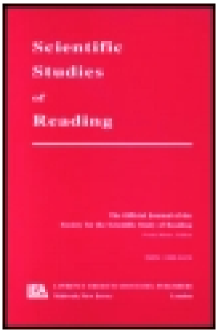

Scientific Studies of Reading
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/hssr20

Effects of Reading Decodable Texts in Supplemental First-Grade Tutoring

Joseph R. Jenkins, Julia A. Peyton, Elizabeth A. Sanders & Patricia F. Vadasy

Available online: 19 Nov 2009

SCIENTIFIC STUDIES OF READING, *8*(1), 53–85

Effects of Reading Decodable Texts in Supplemental First-Grade Tutoring

Joseph R. Jenkins
University of Washington

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

At-risk 1st graders were randomly assigned to tutoring in more or less decodable texts, and instruction in the same phonics program. The more decodable group (n = 39) read storybooks that were consistent with the phonics program. The less decodable group (n = 40) read storybooks written without phonetic control. During the first 30 lessons, storybook decodability was 85% versus 11% for the 2 groups. Tutoring occurred 4 days per week for 25 weeks. A control group did not receive tutoring in phonics or story reading. Both tutored groups significantly surpassed the control on an array of decoding, word reading, passage reading, and comprehension measures. However, the more and less decodable text groups did not differ on any posttest.

Our focus in this article is the decodability of texts used in supplemental tutoring programs for at-risk beginning readers, specifically its effect on the development of word reading skill. Skilled readers can identify most words instantly (by sight), but the ability to read in this manner requires that individuals possess alphabetic reading skill (i.e., decoding) and a history of exposure to the words themselves (Rayner, Foorman, Perfetti, Pesetsky, & Seidenberg, 2001). Alphabetic reading skill is generally measured by performance on nonword (e.g., *stabe*) reading tasks to distinguish it from reading well-learned words from memory.

Whereas the utility of alphabetic reading skill is obvious in learning to read languages with transparent orthographies like Finnish, Italian, and German (Landerl, Wimmer, & Frith, 1997; Paulesu et al., 2001; Wimmer & Mayringer, 2001), it is less so in deep orthographies like English where the associations of phonemes to

Requests for reprints should be sent to Joseph R. Jenkins, College of Education, Box 353600, University of Washington, Seattle, WA 98195–3600. E-mail: jjenkins@u.washington.edu

letter strings are more context sensitive and dependent on morphological structures. However, even the ability to read "exception" words (e.g., *guitar, yacht, sword*) that violate typical grapho-phonemic conventions (i.e., have less overlap with other words) depends on having acquired generalized decoding skill (Gough & Walsh, 1991; Tunmer & Chapman, 1998). Using decoding skill in combination with sentence context, children can identify most regular and exception words on their own. Through such independent learning trials (Share, 1995), children build orthographic memories for words they have encountered.

Although decoding skill is requisite for skilled reading, its acquisition is one of the major hurdles in learning to read. Many children have difficulty detecting, registering, and using the grapho-phonemic regularities inherent in English. Phonics instruction can ease children's reading acquisition (National Reading Panel [NRP], 2000; Snow, Burns, & Griffin, 1998), in part, because it makes explicit the alphabetic principle, draws attention to specific orthographic features of words, teaches the most important grapho-phonemic relationships in the language, and provides strategies (e.g., sound blending) for applying this knowledge (Adams, 1990; Beck & Juel, 1995; Bond & Dykstra, 1967; Chall, 1967; NRP, 2000). Most important, phonics instruction helps children understand that there are no shortcuts to word learning, that all the letters in a word are important (Adams, 1990).

Phonics instruction is thought to produce better results when it is accompanied by ample opportunities to read words that are consistent with the phonics elements targeted. Such practice often occurs as phonics-related word study. At issue is the value of supplementing phonics or word study with practice in decodable text. When it comes to the choice of texts, instructional approaches vary considerably, ranging from texts that emphasize predictability, high-frequency words, children's interests, children's literature, and phonetic-control. Recent research comparing first-grade basal reading programs reveals considerable variation in the texts children read. For example, Foorman, Francis, Davidson, Harm, and Griffin (2002) found large differences among six beginning reading programs in the number of unique words (*types*), total words (*tokens*), word repetitions, words that appear only once (*singletons*), and word decodability and sentence complexity. Hiebert (2002), Hiebert and Fisher (2002), and Stein, Johnson, and Gutlohn (1999) have reported similar findings.

DECODABLE TEXT

Today, one of the loudest controversies in beginning reading instruction centers on the value of decodable texts. On one side of the controversy are claims that beginning readers do not need decodable texts to be successful; The idea that "decodable texts are needed" is "unscientific" (Allington, 1997, p. 15), and "There are many experiments which demonstrate greater reading achievement among early readers working with predictable stories than among early readers working with decodable stories" (Moustafa, 1997, p. 18).

On the other side of the controversy are claims that decodable texts facilitate children's reading acquisition. Foorman, Fletcher, and Francis (1997) wrote,

> there is a period during beginning reading instruction when all children benefit from practicing letter-sound connections in decodable text. To immerse children in a print environment without instruction in letter-sound correspondences and practice in decodable text is to doom a large percentage of children to reading failure. (p. 16)

Beck (1997) acknowledged gaps in the research literature that preclude definitive statements on the necessity and optimal degree of text decodability but suggested that one can reach an "educated conclusion" on the appropriate level of decodability by taking into account principles of learning along with existing research on code-emphasis programs. She wrote, "It would seem that about 70 to 80 percent decodable would be reliable enough for children to refine their knowledge of the spelling-to-speech mapping system, although 30–50 percent is not enough" (p. 17).

Proponents of decodable texts emphasize the critical role of decoding skill in learning to read, the difficulty that many children encounter in acquiring this skill, principles of instruction that should ease its acquisition, and research on the efficacy of code-emphasis reading programs. For example, one of the better researched code-emphasis programs, Distar Reading (Englemann & Bruner, 1978), was designed on the basis of an explicit theory of instruction (Englemann, 1980; Englemann & Carnine, 1996). Following the principle of "introducing examples consistent with a strategy before introducing exceptions" (Carnine, Silbert, & Kame'enui, 1997, p. 12), Distar and other code-emphasis programs employ reading passages "that demonstrate consistent letter-sound correspondences" (p. 12). Emphasizing the connection between phonics and text practice for inchoate readers, Carnine et al. offered guidelines for selecting reading passages:

> During the first weeks of passage-reading, the passages students read should contain only words that have previously appeared in list exercises. (p. 91)
>
> A successful passage-reading component is possible only if the stories presented in the passage-reading exercises are carefully controlled to ensure the student has a strategy to decode every word in the passage. (p. 190)
>
> Most of the passage-reading exercises at that time [until the end of the beginning reading phase] center on decoding practice. The teacher concentrates on providing students with practice in applying decoding skills [first] learned in the word-list exercises. (p. 191)

Juel and Roper-Schneider's (1985) study comes closest to testing the influence of decodable texts on reading development. It compared outcomes for two groups of first graders, both of which received the same phonics instruction but differed in the classroom texts they read. One group used a basal series that "emphasized primarily regular, decodable patterned words in the initial preprimer texts" (p. 139); the other group used a basal series that "focused more on high frequency words and whose text in the initial preprimers exhibited more equality between regular and irregularly decodable patterned words" (p. 139). The researchers did not indicate if the decodable words were consistent with previously taught phonics elements; rather, they equated *decodable* with phonetically regular. Results showed that at the end of the year, the decodable text group had developed significantly stronger decoding skill (ability to read nonwords) and were more able to read words not directly taught by either program. However, the groups did not differ on a norm-referenced reading achievement test (Iowa Test of Basic Skills). Because the children in this research entered first grade with scores above the 40th percentile in reading readiness, generalizations to lower achieving students are limited.

Several other studies with at-risk beginning readers reported strong effects from interventions that combine explicit instruction in phonological awareness, phonics, word study, and reading in decodable texts (Abt Associates, 1977; Blachman, Tangel, Ball, Black, & McGraw, 1999; Foorman, Francis, Fletcher, Schatschneider, & Mehta, 1998; Fuchs et al., 2001; Mathes, Howard, Allen, & Fuchs, 1998; O'Connor, Fulmer, Harty, & Bell, 2001; Slavin, Madden, Karweit, Livermon, & Dolan, 1990; Torgesen, Wagner, Rashotte, et al., 1999; Vadasy, Jenkins, & Pool, 2000). For example, Torgesen, Wagner, Rashotte, et al.'s tutoring study found that a group given explicit instruction in phonics and phonological awareness, including practice in decodable texts, outperformed a less explicit phonics group that did not use decodable texts. Because the two groups differed markedly in the type of phonics tutoring they received, the text effects could not be isolated. Indeed, code-emphasis approaches tend to conflate practice in decodable text with an emphasis on systematic phonics instruction (Beck, 1997), making it impossible to separate the independent contribution of these two variables in most studies. Nevertheless, it is noteworthy that every one of the aforementioned research groups took great pain to match texts with phonics lessons. At the same time, some tutoring studies that did not use decodable text (e.g., Pinnell, Lyons, Deford, Byrk, & Seltzer, 1994; Vellutino et al., 1996) reported significant benefits. None of the aforementioned intervention studies separated the effects of decodable texts from those of other treatment components.

The contribution of decodable texts to reading acquisition remains uncertain and controversial, in part, because researchers have not yet focused on this topic. In their review of phonics instruction, the NRP (2000) named decodable texts as one of three important neglected research topics. This study begins to redress this state of affairs by examining the effects of decodable text practice on learning to read.

The research was conducted with first-grade students at risk for reading failure. Treatment groups received tutoring as a complement to classroom instruction. Tutoring and small group instruction have become a favored secondary-prevention approach for students encountering problems in learning to read (Al Otaiba & Fuchs, 2002; Pinnell et al., 1994; Slavin et al., 1990; Vadasy, Jenkins, & Pool, 2000; Wasik, 1998). Identifying effective components of tutoring programs takes on added importance in the context of current educational efforts to leave no child behind (PL 107-110, No Child Left Behind Act of 2001, 20 U.S.C.A.). In this study, both treatment groups received the same phonics lessons. The more decodable text group also practiced in phonetically controlled texts that were consistent with the phonics lessons, and the less decodable text group practiced in texts that did not control for word decodability. A third group (control) did not receive supplemental tutoring. Thus, the study sought to determine whether decodable texts add value to supplemental reading tutoring for struggling beginning readers.

METHOD

Participants

Students. First-grade teachers from 11 urban public schools identified students in their classes who they considered at risk for reading failure. Of these, 121 met our eligibility criterion, scoring at or below the 25th percentile on the Wide Range Achievement Test–Revised (WRAT–R; Jastak & Wilkinson, 1984) reading subtest. We randomly assigned 95 students to one of two treatment groups (more or less decodable texts) and assigned 26 to a control group. Sample sizes were reduced by attrition during the school year: 17 students moved (6, 6, and 5 from the more decodable, less decodable, and control groups, respectively), 2 students were lost because their parent did not give consent (1 each from the less decodable and control groups), and 3 students were lost because tutors were not available (2 from the more decodable group and 1 from the less decodable group). The final sample consisted of 39 in the more decodable group, 40 in the less decodable group, and 20 in the control, all with a WRAT–R Reading score at or below the 25th percentile (M = 9th percentile). Table 1 shows the sample's gender, ethnicity, and eligibility for special services. The chi-square analyses indicated only one significant group difference; a smaller proportion of students in the control group were White ($p < .05$).

Students were randomly assigned to tutors and treatments within their school, with the following limitation: Tutoring time had to match the classroom reading and language arts schedule, and each classroom teacher had students in both tutoring conditions. Because of the aforementioned attrition, every teacher did not have students in both treatment groups. However, nearly all of the teachers had students in two or more of the three groups.

TABLE 1
Sample Description and Group Comparisons

Classifications	*More Decodable*[a] *n*	%	*Less Decodable*[b] *n*	%	*Control*[c] *n*	%	*Treatment Versus Control* $\chi^2(1)$	*More Versus Less Decodable* $\chi^2(1)$
Sex								
Male	20	51.3	28	70.0	8	40.0	2.799	2.902
Female	19	48.7	12	30.0	12	60.0		
Ethnicity								
White	18	46.2	22	55.0	4	20.0	6.065*	0.618
Non-White	21	53.8	18	45.0	16	80.0		
African American	5	12.8	2	5.0	5	25.0		
Asian	6	15.4	6	15.0	3	15.0		
Hispanic	6	15.4	4	10.0	3	15.0		
Other	4	10.3	6	15.0	5	25.0		
Special Instruction/ Classification								
Limited English Proficiency (LEP/ESL)	14	35.9	9	22.5	4	20.0	0.668	1.717
Title I	24	61.5	21	52.5	12	60.0	0.060	0.658
Special education	1	2.6	2	5.0	0	0.0	0.783	0.321

[a]$n = 39$. [b]$n = 40$. [c]$n = 20$
*$p < .05$.

Tutors. Schools identified 33 paraprofessional tutors. They received scripted phonics lessons, directions for book reading, attendance forms and recording sheets for each student's lesson coverage, and a set of books for reading practice. Research staff provided tutors with 3 hr of formal training in lesson procedures, weekly observations, ongoing coaching in lesson delivery, and monthly follow-up meetings.

Measures

Consistent with recent research in beginning reading instruction (e.g., Foorman et al., 1998; Torgesen, Wagner, Rashotte, et al., 1999), our measures covered a broad range of early reading skill.

Pretests. Assessments addressed receptive vocabulary, letter knowledge, phonological, decoding, and word-reading skills. Receptive vocabulary was measured with the Peabody Picture Vocabulary Test–Revised (PPVT–R; Dunn & Dunn, 1981) which required students to select a picture that best illustrated the meaning of a stimulus word presented orally.

Letter knowledge was assessed using two measures, both of which displayed the uppercase letters of the alphabet in random order. On the naming measure, students were asked to name as many letters as they could in 1 min. On the sounds measure, students were asked to say the sounds for as many letters as they could in 1 min.

Naming speed and phonolological processing skills were assessed with the Comprehensive Test of Phonological Processing (CTOPP; Wagner, Torgesen, & Rashotte, 1999), the Yopp-Singer segmentation test (Yopp, 1988), and a modification of Rosner's (1979) deletion test (modified by Berninger, Thalberg, DeBruyn, & Smith, 1987). Naming speed was assessed on the CTOPP Rapid Letter Naming subtest, a rapid automatic naming measure. This test requires individuals to name as fast as they can five letters that are repeated in random order. The score is the number of seconds required to name all of the letters on two (Forms A and B) 36-item stimulus cards. Ability to accurately repeat nonwords was assessed by the CTOPP Nonword Repetition task. The Yopp-Singer test requires segmentation, with corrective feedback, of 22 orally presented words into their constituent phonemes. The Modified Rosner is a 10-item test requiring deletion of one syllable from a multisyllabic word, half of the items requiring deletion of the initial syllable and half requiring deletion of the final syllable. The score is number correct.

Decoding skill was assessed using the Woodcock Reading Mastery Test–Revised (WRMT–R) Form H Word Attack subtest (Woodcock, 1987), and the Diagnostic Test of Basic Decoding Skills (Bryant, 1975). The latter test requires reading pseudowords that increase in difficulty, until 5 consecutive items are missed or until all 50 items are completed.

Word reading skill was assessed with the WRAT–R Reading subtest and the WRMT–R (Woodcock, 1987, Form H) Word Identification subtest (split-half reliability for first graders is .98). Both tests require reading increasingly difficult words, but the WRAT–R also requires naming 13 uppercase letters and identifying the first 2 letters in the student's name.

Spelling was assessed with the WRAT–R Spelling subtest. This test requires students to copy marks, write their names, and spell dictated words. Two scores were computed for this measure, the raw score total (which incorporates the copied marks, name spelling, and correctly spelled words) and the number of words correctly spelled.

Posttests. All pretest measures were readministered at posttest, with the exception of the PPVT–R. In addition, several assessments were added to the posttest battery. These are described in the following paragraphs.

The Sight Word and Phonemic Decoding subtests of the Test of Word Reading Efficiency (TOWRE; Torgesen, Wagner, & Rashotte, 1999) were added at posttesting. The former requires reading as many words as possible in 45 sec from a list that gradually increases in difficulty. The latter requires reading as many nonwords as possible in 45 sec from a list that increases from 2-phoneme to 10-phoneme nonwords.

Words learned from text reading was measured with a Text Word List of 20 words, 10 from each treatment, none of which had been taught in isolation. Words were drawn from two books that appeared near the highest lessons that the average student had reached (based on records of lesson progress). Words from the more decodable treatment came from *The Fox* (Wright Group, 1999) and *Samantha* (Maslen, 1987). Words from the less decodable treatment came from *The Fantastic Cake* and *The Cooking Pot* (Wright Group, 1999).

To measure reading in context, students orally read from three grade-level passages. One passage each was drawn from two phonetically controlled storybooks, *Mac Gets Well* and *The Goat* (Makar, 1995). Both passages were judged to be highly decodable in relation to the phonic elements taught in the first 50 lessons. The third passage, drawn from *With My Brother* (Houghton Mifflin, 1999), was composed of high-frequency words, and judged to be less decodable based on phonic elements from the first 50 lessons. For each passage, testers recorded the number of words read correctly and the number of errors in 1 min.

Reading comprehension was assessed by WRMT–R Passage Comprehension (Woodcock, 1987). This test requires restoring a missing word from a series of sentences and short passages.

Growth Measures of Word Reading

To measure growth in word reading skill, we constructed three 20-item test forms drawing on the first 300 words of the Fry Instant Word List (Fry, Kress, & Fountoukidis, 2000). For each test, we selected 8 words from the first hundred most frequent words, and 6 words from each of the second and third hundred most frequent words. The three tests were equated for the mean word-frequency (based on Fry et al.'s rank-order of frequency) by selecting test words from three-word blocks, such that the mean frequency for each of the three tests was identical (M = 100.50, SD = 59.74). One of the 20-word test forms was administered in November, another in February, and another in May. Number of words read correctly per test was one dependent variable. In addition, we classified test words as decodable or nondecodable, according to each student's phonics/word study lesson coverage at the time of the test. Thus, the number of decodable words in each test varied for students, according to their previous phonics instruction. This provided a measure of the percentage of decodable words that students read correctly.

Attendance and Lesson Coverage

Tutors filled out an attendance sheet each day, recording the date, lesson number, and storybooks from which students read.

Teacher Reading Instruction Questionnaire

To describe the background (i.e., classroom) reading instruction, classroom teachers completed a questionnaire that indicated their instructional program (e.g., Scott Foresman), specific instructional activities (e.g., stand-alone phonics instruction, integrated phonics instruction, sight-word teaching, use of decodable texts), relative instructional emphases (literature or skills instruction), and estimated minutes of reading instruction. All but two teachers returned the questionnaire. These two teachers account for three students in the study: two students in the less decodable treatment and one student in the control group.

Components of Tutoring Lessons

The phonics/word study component, drawn from Sound Partners (Vadasy, Wayne, et al., 2000), targeted letter sound correspondences (including digraphs), blending letters into words, reading and spelling phonetically regular words, and reading nondecodable and high-frequency words scheduled to appear in the text portion of the lesson. Students also read from storybooks. Lessons were scripted and all tutoring was one-to-one. A component lasted from 2 to 20 min, with storybook reading gradually consuming more time. The components are described in the following sections.

Practicing letter–sound relations. Instruction targeted the most common sound associated with single letters and digraphs. Approximately one new letter–phoneme relation was introduced per lesson, along with a review of previously taught relations. Students also practiced writing graphemes. The sequence of phonic elements is available from the authors.

Reading decodable words. In each lesson, students practiced 8 to 20 words comprising previously introduced letter sounds. Blending letter sounds was taught as a strategy.

Spelling. In this component, students practiced spelling three words selected from the lesson's decodable words. Students were taught to stretch words into their constituent sounds and represent the sounds with written letters, then blend the written product into a spoken word.

Reading nondecodable words. Students practiced nondecodable words (in relation to the phonics taught up to that point) scheduled to appear in their text reading. Often these were high-frequency words. Lessons included new and review words.

Text reading. Each lesson specified the number of minutes for storybook reading, which was the same in both treatments. A new book was introduced every two lessons, starting with Lesson 6. In this component, students read the new book twice, reread the previous book once, then reread any other previous books for the time remaining. Based on students' skill level, the tutor used echo, partner, or independent reading. In echo reading, tutors read a line, with finger-pointing, then students followed by reading the same line. For partner reading, tutor and student read in unison. Time devoted to storybooks increased, beginning with 10 min for Lessons 6 to 30, 15 min for Lessons 31 to 60, and 20 min for Lessons 61 to 100.

If students hesitated for more than 5 sec, or misread a word, the tutor prompted them to use previously taught phonic skills (e.g., isolated specific letters in the word and coached the student to figure out the word), or supplied a letter sound or word, as needed. Following a correction, students read the word, then reread the sentence.

Text Descriptions

More decodable texts. Because there is no consensus on the level of decodability that constitutes a decodable text, we refer to texts as *more decodable* and *less decodable*, letting the percentages of decodable words speak for themselves. In the more decodable storybooks, the majority of words could be read from letter–sound relations and word features taught in the phonics lessons, or had been taught holistically (e.g., nondecodable words). The main book series for the more decodable text group was the Bob Books (Maslen, 1987), supplemented by a few books from Get Ready, Get Set, Read! (Foster, Erickson, & Gifford, 1996) and the Wright Skills set (Wright Group, 1999).

Less decodable texts. These storybooks used fewer words that were decodable from previous phonics instruction. Illustrations provided picture clues, and the books sometimes used repetition to teach new words. All storybooks in this treatment came from the Wright Group: Story Box (2000), Vision Series (1997), and Sunshine Books (1996).

To examine text characteristics in more detail, storybooks were typed into a Microsoft Excel database, from which we developed a list of the unique words in each treatment. Based on the first appearance of each unique word within a treatment, we classified words as decodable or nondecodable. Words were considered decodable if they were comprised of previously tutored phonic elements (i.e., grapho-phonemic correspondences). For example, if only the sounds for *m*, *a*, and *t* had been previously taught, then the words *mat*, *at*, and *am* were considered decodable, but *sat*, and *that* were not, lacking prior instruction in the sounds for *s* and *th*. In determining if a word was decodable, we converted orthographic strings into phonetic codes, such that having taught the long *a* pronunciation for the letter pair *ai*, we classified *mail* and *jail* as decodable, but classified *said* as nondecodable. We also considered words with plural endings (cups-/*s*/; stands-/*z*/; and crashes-/*Iz*/) and past tense markers (chugged-/*d*/; stopped-/*t*/; and started-/*id*/) as decodable only if the ending variants had been taught in a previous phonics lesson. Nondecodable words were those that included one or more grapho-phonemic relations not previously taught. Words with silent letters (e.g., *knock*) were counted as nondecodable. No compound words with morphophonemic shifts, like *vineyard*, appeared in either treatment.

Table 2 shows texts from the 16th lesson of the treatments. The more decodable text has 32 words (i.e., tokens) of which 11 words are unique (i.e., types). By contrast, the less decodable text has 38 tokens, including 13 unique words. For the former, 72.7% of the unique words are decodable, relative to 0% for the latter. If holistically taught words are considered, 100% of the unique words in the more decodable text are readable from previous phonics instruction/word study, versus 16.7% readable unique words in the less decodable text. These texts reveal another

TABLE 2
Sample Texts from More and Less Decodable Groups

Sample Lesson	*More Decodable*	*Less Decodable*
Phonics instruction through Lesson 16		
Consonant sounds	b, c, d, g, h, m, n, r, s, t	b, c, d, g, h, m, n, r, s, t
Vowel sounds	a, i, o	a, i, o
Holistically taught nondecodable words	a, is, his, the	a, see, the, you
Assigned text for Lesson 16		
Assigned book	*Dot and the Dog* (Maslen, 1987)	*Copycat* (Wright Group, 2000)
Text	Dot had *a* bag. *The* bag had *a* tag. *Is the* bag hot? *The* bag *is* hot. Dot had *a* dog. *The* dog *is* Mag. Mag got *the* bag. Dot got Mag.	I go up *the* path. *You* go up *the* path. I go up *the* steps. *You* go up *the* steps. I go into *the* house. *You* go into *the* house. *You* little copycat. I go down *the* steps.
Total unique words in text	11	12
Decodable words	8 (72.7%)	0 (0.0%)
Holistically taught nondecodable words	3 (27.3%)	2 (16.7%)
Readable words	11 (100.0%)	2 (16.7%)

Note. All underlined words judged to be nondecodable based on phonics elements taught in the lessons up to this point; holistically taught nondecodable words are underlined and italicized. Readable words are decodable words plus holistically taught nondecodable words.

characteristic of decodable texts noted by other researchers (Juel & Roper-Schneider, 1985; Menon & Hiebert, 1999); that is, decodable texts often group words with similar spelling patterns. For example, in the more decodable text there are four exemplars each for the short vowels *a* (e.g., *had*) and *o* (e.g., *got*), whereas the less decodable text provides only one exemplar for the vowels *o* and *u* and vowel combinations *ou* and *ow*.

Because we classified a word as decodable or not based on its first appearance, this sometimes resulted in underestimating a word's decodability in later lessons (as more phonics was taught). We partially corrected for this situation by classifying words at three points in the lesson sequence. The 100 lessons were subdivided into thirds. This permitted words classified as nondecodable within the first third of the lessons to be reclassified as decodable in the middle or final third of lessons, if all the pertinent phonic elements had been taught by that time.

Table 3 describes the texts on several dimensions. The description is further subdivided into the first third, the middle third, and the final third of the lesson sequence. The number of unique words (i.e., types) appearing in the two sets of storybooks was 1,201 (more decodable) and 1,110 (less decodable). This is a nonduplicated count (i.e., each unique word is counted only once). To examine word frequencies, we compared the texts for the number and percentage of unique words appearing in the first 300 words of the Fry list of most frequent words (Fry et al., 2000). More and less decodable texts did not differ significantly in this regard. This result held for each of the lesson thirds.

As shown in Table 3, 84.6% of the 91 unique words are decodable in the first third of the more decodable storybooks versus only 10.7% of the 140 unique words in the less decodable storybooks, a practical and statistically significant difference. In the second third, decodability differences between the two sets of storybooks' unique words are still substantial (71.5% vs. 40.2%). In the final third of the lessons, the difference in the percentage of decodable words declines because the percentage of decodable words for the less decodable treatment gradually increases, as more phonic skills are covered during phonics/word study. The decodability difference between treatments is statistically significant at all three points (all ps < .001).

We also examined the number of singletons for each third of the lessons. Singletons are considered an important text dimension because they provide students with little opportunity to commit the word to memory (Hiebert, 2002). In the first third of the lessons, texts in the less decodable treatment have a higher total and a significantly greater proportion of singletons than texts in the more decodable treatment, but thereafter the two texts do not differ significantly in total number or percentage of singletons. Nondecodable singletons are generally considered especially difficult, because they are neither frequent nor readily deciphered. Relative to less decodable texts, the more decodable texts had more singletons that were decodable, and a significantly higher proportion of the unique words was decodable singletons. This result held for the first two thirds of the lessons, but not

TABLE 3
Word Characteristics of Texts

	More Decodable Texts		*Less Decodable Texts*		
Words	*n*	%	*n*	%	$\chi^2(1)$
Texts assigned to first third of lessons					
Total words	510		514		
Total unique words	91		140		
From Fry 300	24	26.4	52	37.1	2.897
Decodable words	77	84.6	15	10.7	125.684***
Nondecodable words	14	15.4	125	89.3	125.684***
Holistically taught nondecodable words	4	4.4	7	5.0	0.044
Singletons	26	28.6	73	52.1	12.513***
Decodable singletons	23	25.3	6	4.3	22.132***
Nondecodable singletons	3	3.3	67	47.9	50.779***
Texts assigned to second third of lessons					
Total words	1,570		1,422		
Total unique words	354		366		
From Fry 300	122	34.5	117	32.0	0.506
Decodable words	253	71.5	147	40.2	71.422***
Nondecodable words	101	28.5	219	59.8	71.444***
Holistically taught nondecodable words	25	7.1	28	7.7	0.091
Singletons	156	44.1	183	50.0	2.542
Decodable singletons	112	31.6	77	21.0	10.444**
Nondecodable singletons	44	12.4	106	29.0	29.821***
Texts assigned to final third of lessons					
Total words	6,069		4,716		
Total unique words	1,036		878		
From Fry 300	212	20.5	208	23.7	2.889
Decodable words	833	80.4	601	68.5	36.146***
Nondecodable words	203	19.6	277	31.5	36.146***
Holistically taught nondecodable words	81	7.8	64	7.3	0.190
Singletons	410	49.6	383	43.6	3.207
Decodable singletons	346	33.4	307	35.0	0.520
Nondecodable singletons	64	6.2	76	8.7	4.306*

Note. The number of unique words for the first, middle, and final thirds are not exclusive (i.e., a unique word counted in the first third of the lessons may appear in a later third). The nonduplicated unique word counts across all storybooks are 1,201 and 1,110 for the more and less decodable treatments, respectively. The comparison of Fry words is conducted for the first 300 Fry word set only. Fry words overlap with other word types (decodable and nondecodable). Singletons are the number of unique words that occur only once in the set of texts described (frequency of the word = 1). Percentages use the number of unique words in the denominator.

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

for the final third. By contrast, relative to the more decodable texts, less decodable text had more nondecodable singletons, and a significantly higher proportion of unique words were nondecodable singletons. This difference held across all three lesson segments (i.e., thirds). In the first third of the lessons, the less decodable texts had more than 20 times the number of nondecodable singletons than did the more decodable texts. In the second third of the lessons, the less decodable texts had more than twice the number of nondecodable singletons than did the more decodable texts. By the final third of the lessons, the difference in the number of nondecodable singletons was less pronounced. The two text types did not differ in the number (or percentage) of holistically taught nondecodable singletons.

To examine sentence complexity for the two types of text, we sampled sentences from three pages of each storybook (using pages taken from the beginning, middle, and end of books), focusing on four sentence characteristics. Table 4 reveals few differences between texts on a per sentence basis for number of words, unique words, prepositional phrases, and clauses. Less decodable texts had more prepositional phrases per sentence in the first third of the lessons, and more clauses per sentence in the final third of the lessons.

Treatment Integrity

To assess treatment integrity, research staff conducted 793 lesson observations, with a mean of 26 observations per tutor and 8 observations per student. Observers used a 27-item criterion checklist reflecting the instructional protocol (including scaffolding) of eight lesson components: letter sounds; sound blending; word reading; spelling; sight-words; sentence reading; specific skill instruction, including word endings and silent-*e* words; and storybook reading. In addition, the checklist included eight items addressing organization, affect, and management of time and materials. For each treatment, we computed the mean percentage of criteria met by tutors on each of the lesson components and the organization/management category. Overall adherence to the instructional protocol was comparable for the two treatment groups, $F(1, 38) = 1.715$, $p > .05$, with most variables exceeding 90%, and no significant between-group differences on any variable.

Rather than assessing interobserver reliability through simultaneous observations by pairs of observers, we examined results of paired observations conducted within a 15-session interval. The latter represents a more stringent estimate of interobserver reliability, as the estimate is subject not only to variation between observers but also to tutor variation across time (somewhat analogous to test–retest reliability using alternate forms). Using observed percentages of treatment integrity for a given tutor–student pair, we divided the smaller percentage by the larger percentage (e.g., for a pair of observations, the first observer's 92% recording of integrity was divided by the second observer's 98% recording). Mean reli-

TABLE 4
Sentence Characteristics of Texts

	More Decodable Texts		*Less Decodable Texts*		
Sentences	*M*	*SD*	*M*	*SD*	*F*
Texts assigned to first third of lessons					
Total words per sentence	4.1	0.90	5.1	1.53	4.062 (1, 24)
Unique words per sentence	3.8	0.96	4.7	1.63	3.256 (1, 24)
Prepositional phrases per sentence	0.2	0.23	0.7	0.62	6.562 (1, 24)*
Clauses per sentence	0.1	0.18	0.2	0.38	0.780 (1, 24)
Texts assigned to second third of lessons					
Total words per sentence	6.9	2.30	6.9	2.81	0.002 (1, 29)
Unique words per sentence	6.2	2.15	6.2	2.76	0.001 (1, 29)
Prepositional phrases per sentence	0.7	0.40	0.6	0.39	0.543 (1, 29)
Clauses per sentence	0.3	0.50	0.3	0.54	0.928 (1, 29)
Texts assigned to final third of lessons					
Total words per sentence	7.4	2.31	8.1	2.38	0.708 (1, 37)
Unique words per sentence	6.7	2.11	6.8	1.76	0.062 (1, 37)
Prepositional phrases per sentence	0.7	0.65	0.5	0.35	2.494 (1, 37)
Clauses per sentence	0.3	0.28	0.6	0.35	7.420 (1, 37)**

Note. Sample data taken from sentences on three pages per book (second page, middle page, and second-to-last page) for texts described. If sentence on sample page begins and/or ends on another page, word count was continued from other applicable pages such that words per sentence were always counted for intact sentences. Unique words are defined as the first appearance of a word within a page sample; word may repeat on other pages of same book and/or other books.

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

ability calculated for 24 tutor–student pairings was .96 (*SD* = 0.08) for lesson components and 1.0 (*SD* =0.00) for management/organization.

Procedure

In September, teachers identified students performing at the bottom of their classes in reading, whom they considered at risk for reading failure. Pretests occurred between mid-September and early October. Students were randomly assigned to more or less decodable treatments. Tutoring began in mid-October and lasted approximately 25 weeks, 4 days a week for 30 min per day. Growth measures of word reading were administered in November, February, and May. Posttests were completed between April and May. Comparable intervals between pre- and posttests were maintained across schools.

RESULTS

We conducted two sets of comparisons: (a) the combined treatment group versus control, and (b) more versus less decodable text groups.

Background Instruction

Table 5 summarizes background (classroom) instruction according to the reading questionnaires. We used student as the unit of analysis because most teachers had students in more than one group. No chi-square comparing treatment and control was significant for percentage of students taught (a) phonics in isolation, (b) phonics integrated with other reading instruction, (c) sight-word instruction, and (d) use of decodable texts. A higher percentage of control students received greater emphasis on skills than literature.

Classroom teachers named their primary reading program and other reading materials they used. Several teachers used more than one reading program. Teachers named Macmillan, Scott Foresman, Ginn, Open Court, and Houghton Mifflin (ordered by use). With one exception, treatment and control conditions did not differ in percentage of students using any reading program; more control (26.3%) than treatment (2.6%) students used Open Court in their classrooms ($p < .001$). By contrast, more and less decodable groups did not differ in use of reading programs. Fortunately, the two treatment students whose teachers reported using Open Court (a program that emphasizes decodable text) were both in the more decodable group.

Analyses of variance (ANOVAs) on the number of minutes teachers reported for (a) classroom reading instruction ($M = 79.0$, $SD = 26.11$ and $M = 77.4$, $SD = 23.59$ for treatment and control groups, respectively), $F(1, 94) = 0.059$, and (b) for explicit teacher-led reading instruction ($M = 26.8$, $SD = 16.83$ and $M = 20.0$, $SD = 14.24$ for treatment and control groups, respectively), $F(1, 94) = 2.594$, revealed no differences between treatment and control groups.

Similarly, none of the chi-square analyses were significant when comparing the two treatment groups on background factors. ANOVAs comparing minutes per day of teacher-reported reading instruction ($M = 79.4$, $SD = 24.63$ and $M = 78.6$, $SD = 27.87$ for more and less decodable groups, respectively), $F(1, 75) = 0.018$, and for minutes of explicit teacher-led reading instruction ($M = 27.6$, $SD = 16.54$ and $M = 25.9$, $SD = 17.31$, for more and less decodable groups, respectively), $F(1, 75) = 0.181$ were not significant.

Amount of Tutoring and Text Reading

Tutored groups (more decodable $M = 89.3$ and $SD = 16.23$; less decodable $M = 95.9$ and $SD = 16.95$) did not differ in the number of tutoring sessions, $F(1, 77) = 3.61$, p

TABLE 5
Background (Classroom) Reading Instruction Characteristics for All Groups

Measures	*More Decodable*[a] *n*	*More Decodable*[a] %	*Less Decodable*[b] *n*	*Less Decodable*[b] %	*Control*[c] *n*	*Control*[c] %	*Treatment Versus Control* $\chi^2(1)$	*More Versus Less Decodable* $\chi^2(1)$
Percent of students whose teachers reported								
Teaching phonics in isolation	30	76.9	28	73.7	17	89.5	1.785	0.109
Integrating phonics in other instruction	14	35.9	12	31.6	8	42.1	0.463	0.160
Teaching sight words	39	100.0	38	100.0	19	100.0	—	—
Using some decodable texts	27	69.2	21	55.3	12	63.2	0.004	1.599
Percent of students whose teachers described their reading approach as								
More emphasis on literature than skills	9	23.1	14	36.8	2	10.5	2.961	1.741
Equal emphasis on literature and skills	21	53.8	16	42.1	7	36.8	0.771	1.063
More emphasis on skills than literature	9	23.1	8	21.1	10	52.6	7.038*	0.046*
Percent of students whose teachers reported using								
Ginn	6	15.4	9	23.7	4	21.1	0.024	0.845
Houghton Mifflin	2	5.1	1	2.6	3	15.8	3.679	0.320
MacMillan/McGraw-Hill	20	51.3	18	47.4	5	26.3	3.270	0.118
Open Court	2	5.1	0	0.0	5	26.3	12.682**	2.001**
Scott Foresman	13	33.3	17	44.7	5	26.3	1.052	1.052

Note. Of 23 teachers, 1 teacher with three treatment students (one more decodable and two less decodable) and one control student reported using books from three reading basals (Ginn, MacMillan/McGraw-Hill, and Scott Foresman); and one teacher with five treatment students (two more decodable and three less decodable) and one control student reported using books from two series (Ginn and Scott Foresman).

[a]$n = 39$. [b]$n = 38$. [c]$n = 19$.

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

$> .05$, or the highest lesson number completed (more decodable $M = 78.8$ and $SD = 17.73$; less decodable $M = 75.5$ and $SD = 17.21$), $F(1, 77) = 0.764$, $p > .05$. Every student completed the first third of the tutoring lessons, with most also completing the second third of the lessons (89.7% and 80.0% for the more and less decodable groups, respectively). No student completed all 100 lessons, but 20.5% (more decodable) and 22.5% (less decodable) completed 80 lessons.

We also estimated the number of words students read in storybooks, using the number of lessons a student completed and the total number of words in books students read through their highest lesson. This procedure underestimates the amount of reading performed because minutes devoted to text reading was controlled for each lesson, and students read all books more than once. Nevertheless, using this very conservative estimate of words read during the text reading component, the more decodable group read 8,525.64 words ($SD = 4{,}749.38$), and the less decodable group read 7,092.25 words ($SD = 3{,}529.85$), $F(1, 79) = 2.326$, $p > .05$. The estimated mean number of unique words read during text reading was 694.7 ($SD = 322.1$), and 670.3 ($SD = 254.2$) for more and less decodable groups, respectively, $F(1, 79) < 1$, $p > .05$.

Pretests

ANOVAs comparing pretests for treatment versus control group and for more versus less decodable text groups were not significant for any measure (Table 6).

Posttests—Treatment Versus Control Groups

Multivariate analyses of variance or covariance (MANOVAs or MANCOVAs) on conceptually similar posttests were computed with a relevant pretest as a covariate. The WRMT–R Word Attack pretest was the covariate for decoding posttests. The WRMT–R Word Identification pretest was the covariate for word reading posttests. PPVT–R pretest was the covariate for reading comprehension. WRAT–R Spelling pretest was the covariate for spelling measures. Adjusted means, standard deviations, and effect sizes are shown in Table 7.

Decoding. The MANCOVA for decoding posttests was significant. Univariate tests revealed the treatment group significantly outperformed controls on Bryant's (1975) decoding test and WRMT–R Word Attack, but the groups did not differ reliably on the TOWRE Phonemic Decoding posttest.

Word reading and spelling. The MANCOVA for word reading was significant, favoring the treatment over the control group. Univariate *F*s were significant

TABLE 6
Means, Standard Deviations, and Group Comparisons on Pretests

Measures	*More Decodable*[a]		*Less Decodable*[b]		*Control*[c]		*Treatment Versus Control*	*More Versus Less Decodable*
	M	*SD*	*M*	*SD*	*M*	*SD*	*F*	*F*
Age (years)	6.55	0.39	6.65	0.36	6.55	0.27	0.406 (1, 97)	1.324 (1, 77)
Receptive language comprehension								
PPVT–R	89.67	20.16	90.18	19.39	90.05	18.75	0.001 (1, 96)	0.013 (1, 76)
Letter knowledge								
Letter naming (letters per minute)	34.31	12.58	36.03	14.10	34.65	11.32	0.026 (1, 97)	0.326 (1, 77)
Letter sounds (letters per minute)	11.59	7.98	11.85	9.22	14.35	8.06	1.534 (1, 97)	0.018 (1, 77)
Phonological knowledge								
CTOPP Rapid Letter Naming (seconds)	113.46	49.57	112.30	44.33	114.53	28.89	0.022 (1, 97)	0.012(1, 77)
Yopp-Singer	3.72	5.20	5.35	6.40	5.63	5.76	0.531 (1, 96)	1.543 (1, 77)
Modified Rosner	6.31	3.07	6.58	3.08	6.11	3.02	0.188 (1, 96)	0.149 (1, 77)
CTOPP Nonword Repetition	7.23	2.92	6.68	2.79	7.58	3.34	0.698 (1, 96)	0.748 (1, 77)
Decoding								
WRMT–R Word Attack	0.38	1.46	0.13	0.56	0.10	0.31	0.375 (1, 97)	1.095 (1, 77)
Word reading								
WRMT–R Word Identification	2.13	2.98	2.33	3.83	1.80	1.70	0.294 (1, 97)	0.065 (1, 77)
WRAT–R Reading	25.41	3.56	25.70	3.38	26.35	1.95	0.972 (1, 97)	0.138 (1, 77)
Spelling								
WRAT–R Spelling	18.33	2.65	18.93	3.21	17.55	5.87	1.364 (1, 97)	0.795 (1, 77)
WRAT–R spelling words correct	0.92	0.98	1.08	1.37	2.10	5.01	3.187 (1, 97)	0.320 (1, 77)

Note. All raw scores except PPVT–R (where standard scores were used). All *p*s > .05. PPVT–R = Peabody Picture Vocabulary Test–Revised; CTOPP = Comprehensive Test of Phonological Processing; WRMT–R = Woodcock Reading Mastery Test–Revised; WRAT–R = Wide Range Achievement Test–Revised.

[a]$n = 39$. [b]$n = 40$. [c]$n = 20$.

TABLE 7
Adjusted Means, Standard Deviations, and Group Comparisons on Posttests

	More Decodable[a]		*Less Decodable*[b]		*Control*[c]		*Treatment Versus Control*		*More Versus Less Decodable*	
Measures	*M*	*SD*	*M*	*SD*	*M*	*SD*	*F*	*d*[d]	*F*	*d*[d]
Decoding[e]							8.109 (3, 94)***		1.137 (3, 74)	
Bryant test	20.14	10.96	21.49	10.96	9.42	10.12	20.247 (1, 96)***	1.13	0.294 (1, 76)	–.12
WRMT–R Word Attack	15.49	8.75	13.92	8.75	8.27	8.39	9.325 (1, 96)**	.77	0.626 (1, 76)	.18
TOWRE Phonemic Decoding	10.72	7.70	10.75	7.70	8.04	7.18	2.256 (1, 96)	.38	0.000 (1, 76)	.00
Word Reading[e]							4.761 (4, 93)**		0.352 (4, 73)	
WRMT–R Word Identification	32.98	12.54	32.70	12.54	26.72	11.99	3.976 (1, 96)*	.50	0.010 (1, 76)	.02
WRAT–R Reading	47.00	8.62	46.55	8.62	40.66	8.20	8.684 (1, 96)**	.74	0.055 (1, 76)	.05
TOWRE Sight Word Reading	27.61	11.00	26.76	11.00	21.49	10.72	4.336 (1, 96)*	.52	0.117 (1, 76)	.08
Text word list	11.63	5.02	10.94	5.02	6.93	4.95	12.134 (1, 96)***	.87	0.379 (1, 76)	.14
Growth Measure[f,g] (total words read)									1.745 (3, 75)	
November	4.18	2.38	4.80	2.48	—	—	—	—	1.283 (1, 77)	–.26
February	12.67	4.18	11.95	4.4	—	—	—	—	0.550 (1, 77)	.17
May	15.87	3.84	15.23	4.73	—	—	—	—	0.444 (1, 77)	.15
Reading comprehension[h]										
WRMT–R Passage Comprehension	14.97	5.95	14.35	5.95	9.76	6.09	10.354 (1, 96)**	.81	0.213 (1, 76)	.10

Phonetically controlled passages[g]							3.156 (2, 96)*		0.134 (2, 76)	
Fluency (words correct per min)	41.72	25.96	40.90	29.08	27.70	22.03	4.225 (1, 97)*	.52	0.017 (1, 77)	.03
Accuracy (percent correct)	0.82	0.13	0.80	0.19	0.71	0.14	6.093 (1, 97)*	.62	0.214 (1, 77)	.10
Nonphonetically controlled passages[g]							2.013 (2, 96)		0.061 (2, 76)	
Fluency (words correct per min)	35.26	23.11	36.98	25.09	26.35	17.70	2.909 (1, 97)	.43	0.100 (1, 77)	–.07
Accuracy (percent correct)	0.80	0.15	0.81	0.20	0.73	0.17	3.563 (1, 97)	.47	0.098 (1, 77)	–.07
Spelling[e,g]							4.174 (2, 95)*		0.191 (2, 75)	
WRAT–R Spelling	28.46	4.08	28.73	4.09	26.35	4.03	4.661 (1, 96)*	.54	0.090 (1, 76)	–.07
WRAT–R spelling words correct	9.74	3.65	9.78	3.64	7.27	3.52	7.723 (1, 96)**	.70	0.001 (1, 76)	–.01

Note. WRMT–R = Woodcock Reading Mastery Test–Revised; TOWRE = Test of Word Reading Efficiency; WRAT–R = Wide Range Achievement Test–Revised. All raw scores.

[a]$n = 39$. [b]$n = 40$. [c]$n = 20$. [d] Effect size calculated as the difference in adjusted group means divided by the square root of the Mean Square error. [e]Multivariate analysis of covariance Wilks's Lambda. Decoding uses WRMT–R Word Attack pretest as covariate; Word Reading uses WRMT–R Word Identification pretest as covariate; Spelling uses WRAT–R Spelling pretest as covariate. [f]Growth measure is a selection of 20 words, randomly ordered for each test interval, taken from the first 200 high-frequency words of the Fry word list (Fry et al., 2000). A new set of words was used for each test interval, such that mean word frequency of the tests was identical at each test interval, $M = 100.50$, $SD = 59.74$, $F(2, 57) = 0.000$, $p = 1.00$. [g]Multivariate analysis of variance Wilks's Lambda. [h]Analysis of covariance; Peabody Picture–Vocabulary Test–Revised used as pretest covariate.

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

for all single word reading measures—WRAT–R Reading, WRMT–R Word Identification, TOWRE Sight Word, and Text Word list. The spelling MANCOVA and both univariate tests significantly favored the tutored groups.

Reading comprehension and reading in context. Tutored groups significantly outperformed controls on WRMT–R Passage Comprehension. In assessing reading accuracy and fluency in context, scores from the phonetically controlled and phonetically uncontrolled passages were analyzed separately. The MANOVA and univariate tests for the phonetically controlled passages showed tutored students significantly outperforming controls (i.e., a 14-word reading fluency and 11-point accuracy advantage). By contrast, the MANOVA on the phonetically uncontrolled passage was not significant, although the tutored group showed a 10-word advantage in reading fluency and a 9-point advantage in accuracy. The different results for the two types of passages may stem from the specific passages employed (rather than phonetic control) or they may indicate that tutoring in phonics and word study resulted in specific effects on more decodable texts. Lacking a norm reference for these passages, it is difficult to evaluate students' accuracy and fluency levels of oral reading. Nevertheless, the combined treatment group's oral reading performance in context was low by most standards (Hasbrouk & Tindal, 1992; Stahl & Heubach, in press). This contrasts with single word reading levels above the 50th percentile (mean standard scores of the treatment group on the WRMT–R word identification and word attack were 104 and 108, respectively).

Posttests—More Versus Less Decodable Groups

No MANOVAs, MANCOVAs, or univariate tests comparing more and less decodable treatments were significant on any outcome (see Table 7).

Results for the lowest performing students. The majority of students in both treatments attained grade-level on WRMT–R Word ID (84.6% and 77.5% for the more and less decodable groups, respectively) and Word Attack (87.2% and 90.0% for the more and less decodable groups, respectively). We defined grade-level performance as scoring no more than one standard error of measurement below the 50th percentile. However, not all students demonstrated desirable achievement levels. We defined treatment nonresponders as students scoring at or below the 25th percentile. Using this criterion with WRMT–R Word Identification, 10.3% of the more decodable and 15.0% of the less decodable treatment were nonresponders; on the WRMT–R Word Attack 3.0% and 7.5% of the more and less decodable groups, respectively, were nonresponders. Neither difference was statistically significant, both $\chi^2(1) < 1$.

Using a two-step procedure, we examined a subgroup deemed most likely to encounter problems learning to read. In Step 1, we created a composite phonological-letter knowledge score for each student by combining *z* scores for the phonological (Modified Rosner syllable deletion, Yopp-Singer Segmentation, CTOPP Rapid Letter Naming, and CTOPP Nonsense Word Repetition) and letter knowledge (Names and Sounds) pretests. In Step 2, we identified students with below median composites, yielding 19 students in each of the two treatments. ANOVAs comparing groups on Word Identification and Decoding posttests revealed no significant differences, with all Fs(1, 36) < 1: WRMT–R Word ID, $d = 0.14$; WRAT–R Reading, $d = 0.11$; WRMT–R Word Attack, $d = 0.02$; Bryant, $d = 0.25$. A similar analysis was done on an even more extreme subgroup—students whose composite phonological processing/letter knowledge *z* score placed them in the bottom quartile of the at-risk sample ($n = 9$ and 10 for more and less decodable groups, respectively). Again, no significant differences were found between treatment groups on the four posttests of interest, with all F's(1, 17) < 1.0: WRMT–R Word Identification, $d = 0.21$; WRAT–R Reading, $d = 0.06$; WRMT–R Word Attack, $d = 0.03$; Bryant, $d = 0.26$.

Growth Curve Analyses

More and less decodable groups' growth in word learning (three probes of 20 high-frequency words) was examined with Hierarchical Linear Modeling (HLM) growth curve analysis. Table 8 shows the results of the fixed effect growth curve model for the text factor and the random effects analysis for individual variability in word reading growth. Both groups progressed significantly in word reading accuracy, but differential effects for treatment group did not materialize (Figure 1). The random-effects analysis showed no significant variability in individual growth. Together, the fixed effect growth curve analysis and random effects analysis indicate that all tutored students made significant growth, without a differential effect due to type of text practice or individual variation about the groups' growth. Students who read relatively fewer words correctly at the beginning of the year ended the year reading fewer words, and students who began the year reading relatively more words ended the year reading more words. In a second growth curve analysis, we used percentage of correctly read decodable words as the dependent variable. Because results from the decodable-word analysis were consistent with the total-word analysis, we table only the latter. In summary, results of growth curve analyses were consistent with results of posttests (i.e., no reliable differences between treatments).

TABLE 8
Fixed and Random Effects of Growth Measure by Treatment

Fixed Effects	*Coefficient*	*SE*	*t*
Words Read – Initial Status	5.26	0.33	15.940[a] *
Words Read – Slope	5.53	0.23	24.040[b] *
Contribution of Treatment	–0.59	0.45	–1.295[b]
Random Effects	*SD*	*Variance Component*	χ^2
Words Read – Initial Status	1.87	3.51	114.227[a] *
Slope	1.06	1.13	86.374[b]
Error	2.47	6.09	

[a]*df* = 78. [b]*df* = 77.
**p* < .01.

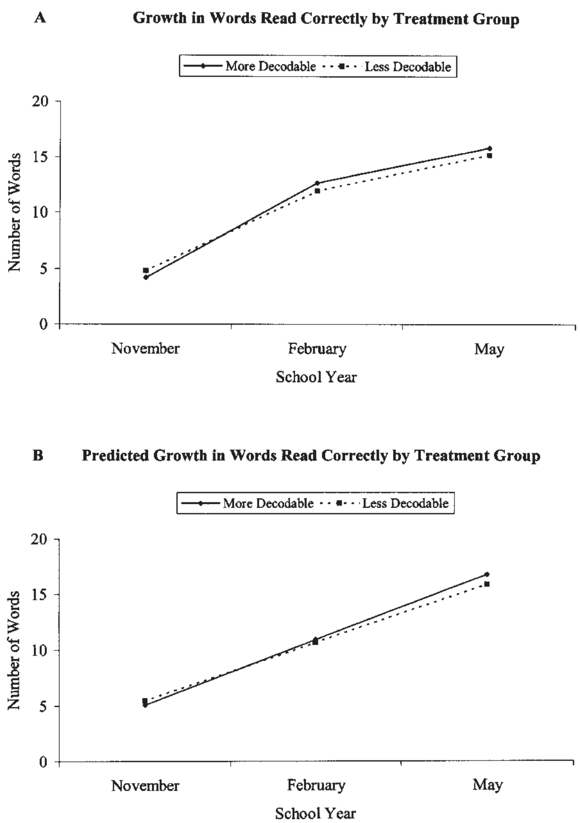

FIGURE 1 Growth in word reading raw scores by treatment (Panel A) and predicted growth in word reading scores by treatment (Panel B).

DISCUSSION

Our primary interest was whether tutoring in decodable texts facilitated reading achievement of at-risk first graders. Results showed broad effects from tutoring in phonics/word study and storybook reading, but no reliable effects for differing levels of text decodability.

Eight study features lend weight to the results. First, sample sizes for the two treatment groups (39 and 40) were relatively large for yearlong tutoring studies. Second, students were drawn from 23 first-grade classrooms from 11 schools, and were taught by 40 different tutors. Third, the background (classroom) reading programs of treatment groups were similar. Fourth, students were randomly assigned to the two treatments, and were comparable in gender, age, and pretest achievement. Fifth, both treatment groups received identical phonics instruction, differing only in the level of decodability of their practice texts. Sixth, the amount of text reading that occurred in the treatment was considerable, with the most conservative estimates indicating 7100 to 8500 running words read. Seventh, group differences were consistent across a variety of reading outcomes (decoding, word reading, spelling, reading in context, and comprehension), with multiple measures employed on most outcomes. Finally, treatment integrity, rarely measured in intervention research (Al Otaiba & Fuchs, 2002), was high for both treatment groups.

At the same time, four study features limit generalization. First, students were not randomly assigned to the treatment–control comparison. Although schools were agreeable to randomly assigning students to one of two treatments, we were unsuccessful in influencing their decisions to tutor or not tutor students. Control and treatment students were drawn from similar classrooms in the same school district and were similar on most demographic and background instructional variables. However, significantly more control students were not White (although groups did not differ in English Language Learners or special education status). Treatment and control groups scored at similar levels on all pretests, with the control group performing higher on 10 of 12 pretests, although none of the differences approached significance.

Tutored and control students had a similar proportion of teachers reporting classroom instruction in phonics (stand-alone and integrated), sight-words, and reading from decodable texts, and had similar amounts of time devoted to both overall classroom reading instruction and teacher-led reading instruction. However, two differences in background instruction were noted; a significantly higher proportion of control students had teachers who reported (a) using Open Court Reading, and (b) placing greater emphasis on skills, particularly word recognition, relative to children's literature. Had the differences between treatment and control groups' background instruction been in the opposite direction (i.e., less emphasis on word-level skills in control classrooms, less use of code-based instruction), our confidence in attributing treatment effects to tutorial instruction (rather than to differences in background instruction) would weaken. If anything, these background differences between treatment and control groups should favor the control group on word-level reading outcomes. Two other factors are important in gauging whether tutoring contributed to reading growth. First, findings from the treatment–control group comparison are in line with those from three previous studies of this same intervention (Vadasy, Jenkins, Antil, Wayne, & O'Connor, 1997a, 1997b; Vadasy, Jenkins, & Pool, 2000). Second, posttests levels of the treatment group were near the 50th percentile in decoding and word identification, compared to 9th percentile pretests. Although random assignment to the treatment–control comparison would have strengthened the study, it is uncommon in most school-based research. More important, random assignment to the two treatments occurred for the central research question (i.e., levels of text decodability).

A second design feature that limits the breadth of generalization is the context of the intervention. The treatment comparison of primary interest (i.e., reading practice in more or less decodable texts) was situated in tutorial instruction. Relative to classroom instruction, tutorials allow for greater instructional scaffolding. When tutees encounter difficulty in reading a word, tutors can provide graduated assistance (e.g., isolate known letters, prompt use of phonics knowledge, assist in sounding-out, remind when a word is one that has been practiced earlier, or supply the word when lesser prompts do not succeed). Treatment integrity observations indicated that our tutors followed this very protocol. Such scaffolding may have reduced the effect of text decodability on reading acquisition by providing students in both treatments with a sufficient number of opportunities to apply phonics knowledge and learn words.

Moreover, tutorial reading practice in more or less decodable storybooks constituted just one part of students' reading experiences. Students read from a variety of texts in their classrooms, some more decodable than others, with words bearing an unknown relation to the phonic elements taught during tutoring. Although 60% of the teachers reported using some decodable texts in their classrooms, only two reported using a code-emphasis reading program (i.e., Open Court), neither of whom taught students in the treatment groups. Nevertheless, the mixture of texts across classroom and tutorial settings dilutes text-decodability differences between the two treatment groups. This does not mean that text effects within tutorial programs will always be swamped by variations in complementary classroom texts, as O'Connor et al. (2002) recently showed. Our findings in this study leave open the question of whether different levels of text decodability in classroom instruction affect reading achievement. The different findings from Juel and Roper-Schneider (1985) and this study could be due to differences in the two study designs (e.g., classroom instruction vs. supplementary tutoring, students' entering ability, different texts and phonics programs).

Nonetheless, the instructional situation in this study resembles that in many Title 1 and special education programs where student receive one instructional approach in their classrooms, and a different instructional approach outside their classrooms (Johnston, Allington, & Afflerbach, 1985). Government statistics indicate that pull-out programs are a prominent part of the education landscape: 74% of elementary schools offered pull-out Chapter 1 services in 1992 (U.S. Department of Education, 1996) and 83% of students with learning disabilities received pull-out instruction in the 1997–98 year (U.S. Department of Education, 2000). It is to these supplementary programs that our results are most relevant.

Third, our treatments consisted of specific storybooks and a particular phonics program, both of which limit generalization to tutoring programs using different texts and different phonics programs. Although the particular phonics program employed in this study has a strong record of improving reading outcomes, one cannot assume that phonics teaching and phonics learning are isomorphic (Hiebert, 2002). At the same time, the NRP (2000) found no evidence that specific phonics programs and approaches were differentially effective.

Fourth, we classified words as decodable according to one definition of decodability. The meaning of decodable can vary depending on the standard used. Like Barr and Dreeban (1983), Beck (1981), Foorman et al. (2002), and Stein et al. (1999) we used a "lesson-to-text-match" (Mesmer, in press) standard for determining decodability (i.e., words were considered decodable if the relevant grapheme–phoneme elements had been previously taught). However, lesson-to-text-match is not the only standard for defining decodability. Some researchers have used regularity of spelling and phonics patterns as the standard, independent of phonics instruction (Hoffman, Roser, Salas, Patterson, & Pennington, 2000; Juel & Roper-Schneider, 1985; Menon & Hiebert, 1999). For example, Juel and Roper-Schneider used bigram versatility (e.g., the number of words in which pairs of letters co-occur in the same position, as the *ee* in *seed* and *sleep*) to gauge decodability. In addition, judgments of decodability can be indexed dichotomously or continuously. That is, a word can be classified as decodable or nondecodable (our approach) or rated for degree (0–100%) of decodability (C. A. Perfetti, personal communication, July 18, 2002). For example, *shift* is 75% decodable using a lesson-to-text-match, if the *i*, *f*, and *t* but not *sh* have been taught. Such definitional variations yield different text descriptions and should be investigated.

Finally, decodable texts may exert more influence in the earliest stages of reading acquisition. With the exception of the growth measure, our assessments were concentrated at year's end, possibly too late to detect text effects. However, if the benefits of practice in decodable text disappear by the end of first grade, they are short lasting.

Breadth of Effects

The treatment comparison focused on at-risk first graders who were predicted to struggle in learning to read and who, as a group, performed at the 9th percentile on a word reading pretest. Teacher-judgment and pretest achievement predict first-grade outcomes reasonably well, but ascertaining true reading disability in beginning first graders is an imperfect science (Jenkins & O'Connor, 2002; O'Connor & Jenkins, 1999). Thus, it is fair to ask whether our results can be generalized to children with more severe limitations in developing print–speech codes. Several analyses suggest that findings for the larger at-risk group apply equally well to a lower performing subgroup of the study sample. In an HLM analysis modeling reading growth across the school year, text decodability did not interact with student ability to influence growth in word reading. Similar results were obtained in a more nuanced growth analysis that considered only decodable words (defined on the basis of students' individual lesson coverage at the time of testing). According to this analysis, the two text groups developed decoding skill on a comparable time course. In another analysis, we examined word reading and decoding posttests for a subset of the most at-risk students. These were students whose composite score on phonological processing and letter knowledge pretests placed them in the bottom half of the at-risk group. Level of text decodability did not differentially affect these students' performance on any of the four measures examined. Similar findings were obtained for an even more challenged group (those whose phonological processing/letter knowledge composite score placed them in the bottom quartile of the at-risk group). Finally, although most tutored students achieved grade-level performance on the WRMT–R Word Identification and Word Attack subtests, not every student reached this standard. Comparable percentages of students in the more and less decodable treatments ended the year reading below the 25th percentile. Taken together, these analyses indicate that regardless of students' entering abilities or response to treatment, no effects emerged for text decodability.

CONCLUSIONS

This study adds methodologically and substantively to the literature on the effects of text decodability. With respect to methodology, it included observational data on treatment integrity and employed random assignment for the principal treatment comparison, unlike much of the school-based research in beginning reading (e.g., Abt Associates, 1977; Blachman et al., 1999; Foorman, et al., 1998; Hiebert, Colt, Catto & Gary, 1992; Iverson & Tunmer, 1993; Juel & Roper-Schneider, 1985; Pinnel et al., 1994; Slavin et al., 1990). Regarding experimental designs, we should note that the only other study investigating the effects of decodable text (Juel & Roper-Schneider, 1985) used a quasi-experimental design. With respect to substance, this study focused on at-risk first grade readers, whereas Juel and Roper-Schneider examined students performing above the 40th percentile on a reading readiness test. In addition, we investigated decodable text effects in the context of a supplementary reading intervention, as opposed to classroom reading instruction. This is an important addition because current reading reform efforts focus on both primary- (i.e., classroom) and secondary-level interventions (in which struggling students receive supplementary instruction). In school-based intervention programs, decisions about using decodable text must be made in contexts much like those in our study, where classroom teachers employ a variety of reading programs and texts. It will be important for future studies to seek replication across a variety of texts, phonics programs, and other situational factors, as well as to examine interactions stemming from the combination of specific classroom texts and the texts used in secondary-level interventions.

The strongest generalizations of our results might go something like this. Supplemental phonics instruction along with successful practice in text reading (regardless of decodability levels of the texts) may be sufficient for a majority of at-risk first graders to reach grade-level in two critical areas of reading—development of word-specific representations in memory and skill in decoding unfamiliar words. Decodable texts do not add value to supplemental tutoring programs, even for students who demonstrate more serious limitations in acquiring print–speech codes. However, accepting these generalizations overlooks an important fact; although effect sizes for the text variable were exceedingly small, our result is essentially a nondifference. Generalization is also limited by the specifics of our research design (a particular phonics program, a particular set of controlled storybooks, highly supportive tutorials, and classroom reading instruction in a range of texts unrelated to those used in the treatment).

Rather, our findings indicate that substantial differences in text decodability delivered in tutorials do not always “power through” other classroom and text factors. Instructional and text variables (e.g., the number of word types and tokens, repetitions, singletons, word frequency, and sentence complexity) residing in the background or the tutorial programs may swamp differences in the decodability of texts used in tutoring. Even nominal categories such as “code-emphasis reading programs” may be misleading when text factors are scrutinized. Of the six first-grade programs analyzed by Foorman et al. (2002), the two most decodable programs introduced both the most and the least number of words (types). Amount and type of phonics instruction may also interact with the effects of decodable text. Further complicating matters, these various text and instructional factors play out at two levels for struggling readers (i.e., classroom instruction and supplemental reading instruction). Thus, this study (or any other, for that matter) with its inherent particularities cannot, by itself, settle an instructional issue as complex as text decodability.

ACKNOWLEDGMENTS

This research was supported in part by Grant No. H324C980039 from the U.S. Department of Education, Office of Special Education Programs. Statements do not reflect official policy of any agency.

We thank the participating children and tutors at our research sites. We acknowledge the collaboration of school staff in carrying out this research and their willingness to participate in our longitudinal studies of reading interventions for at-risk students. We also thank Joe Hankins for help on the word databases, Heidi Brough for administrative assistance, and Scott Stage for consultation on HLM analyses.

REFERENCES

Abt Associates. (1977). *Education as experimentation: A planned variation model (Vol. IV-B, Effects of Follow Through Models)*. Cambridge, MA: Abt Books.

Adams, M. J. (1990). *Beginning to read: Thinking and learning about print.* Cambridge, MA: MIT Press.

Allington, R. L. (1997, August/September). Overselling phonics. *Reading Today*, p. 15.

Al Otaiba, S., & Fuchs, D. (2002). Characteristics of children who are unresponsive to early literacy instruction: A review of the literature. *Remedial and Special Education, 23*, 300–316.

Barr, R., & Dreebman, R. (1983). *How schools work.* Chicago: University of Chicago Press.

Beck, I. L. (1981). Reading problems and instructional practices. In G. A. MacKinnon & T. G. Waller (Eds.) *Reading research: Advances in theory and practice, Vol. 2* (pp. 53–93). New York: Academic.

Beck, I. L. (1997, October/November). Response to "Overselling phonics." *Reading Today,* p. 17.

Beck, I., & Juel, C. (1995). The role of decoding in learning to read. *The American Educator, 8*, 21–25, 39–42.

Berninger, V. W., Thalberg, S., DeBruyn, I., & Smith, R. (1987). Preventing reading disabilities by assessing and remediating phonemic skills. *School Psychology Review*, 16, 554–565.

Blachman, B. A., Tangel, D. M., Ball, E. W., Black, R., & McGraw, C. K. (1999). Developing phonological awareness and word recognition skills: A two-year intervention with low-income, inner-city children. *Reading and Writing: An Interdisciplinary Journal, 11*, 239–273.

Bond, G., & Dykstra, R. (1967). The cooperative research program in first-grade reading instruction. *Reading Research Quarterly, 2,* 5–142.

Bryant, N. D. (1975). *Diagnostic test of basic decoding skills.* New York: Columbia.

Carnine, D. W., Silbert, J., & Kameenui, E. J. (1997). *Direct instruction reading*. Upper Saddle River, NJ: Prentice Hall.

Chall, J. S. (1967). *Learning to read: The great debate*. New York: McGraw-Hill.

Dunn, L. M., & Dunn, L. (1981). *Peabody Picture Vocabulary Test–Revised*. Circle Pines, MN: American Guidance Service.

Engelmann, S. (1980). *Direct instruction*. Englewood Cliffs, NJ: Prentice-Hall.

Engelmann, S., & Bruner, E (1978). *Reading Mastery 1*. Chicago: Science Research Associates.

Engelmann, S., & Carnine, D. (1996). *Theory of instruction: Principles and applications*. Eugene, OR: Association for Direct Instruction.

Foorman, B. R., Fletcher, J., & Francis, D. J. (1997, October/November). "Do children understand what they're reading?": Yes, when they've been taught to read. *Reading Today*, p. 16.

Foorman, B. R., Francis, D. J., Davidson, K. C., Harm, M. W., & Griffin, J. (2002, February). *Variability in Text Features in six grade 1 basal reading programs.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Foorman, B. R., Francis, D. J., Fletcher, J. M., Schatschneider, C., & Mehta, P. (1998). The role of instruction in learning to read: Preventing reading failure in at-risk children. *Journal of Educational Psychology, 90*, 37–55.

Foster, K. C., Erickson, G. C., & Gifford, K. (1996). *Barron's books*. Hauppauge, NY: Barron's Educational Series.

Fry, E. B., Kress, J. E., & Fountoukidis, D. L. (2000). *The Reading Teacher's book of lists.* Upper Saddle River, NJ: Prentice Hall.

Fuchs, D., Fuchs, L. S., Thompson, A., Al Otaiba, S., Yen, L., Yang, N., et al. (2001). Is reading important in reading-readiness programs? A randomized field trial. *Journal of Educational Psychology, 93*, 251–267.

Gough, P. D., & Walsh, M. (1991). Chinese, Phoenicians, and the orthographic cipher of English. In S. Brady & D. Shankweiler (Eds.), *Phonological processes in literacy* (pp. 199–209). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Hasbrouk, J. E,. & Tindal, G. (1992). Curriculum-based oral frequency norms for students' in grades 2 through 5. *Teaching Exceptional Children, 24*, 41–44.

Hiebert, E. H. (2002, February). *Describing the Tasks of Texts for Beginning Readers*. Paper presented at the Pacific Coast Research Conference, La Jolla, CA.

Hiebert, E. H., Colt, J. M., Catto, S. L., & Gury, E. C. (1992). Reading and writing of first-grade students in a restructured Chapter 1 program. *American Educational Research Journal, 29*, 545–572.

Heibert, E. H.,& Fisher, C. W. (2002, April). *The critical word factor in texts for beginning readers: Effects on reading speed, accuracy, and comprehension.* Paper presented at annual meeting of the American Educational Research Association, New Orleans, LA.

Hoffman, J. V., Roser, N. L., Salas, R, Patterson, E., & Pennington, J. (2000). *Text leveling and little books in first-grade reading* (CIERA Report # 1–010). Ann Arbor, MI: Center for the Improvement of Early Reading Achievement.

Houghton Mifflin. (1999). *Invitations to literacy*. Boston: Author.

Iversen, S., & Tunmer, W. E. (1993). Phonological processing skills and reading recovery program. *Journal of Educational Psychology, 85*, 112–126.

Jastak, S., & Wilkinson, G. S. (1984). *The Wide Range Achievement Test-Revised.* Wilmington, DE: Jastak Associates.

Jenkins, J. R., & O'Connor, R. E. (2002). Early identification and intervention for young children with reading/learning disabilities. In R. E. Bradley, L. Davidson, & D. P. Hallahan (Eds.), *Identification of learning disabilities: Research to practice* (pp. 99–150). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Johnston, P., Allington, R., & Afflerbach, P. (1985). The congruence of classroom and remedial reading instruction. *Elementary School Journal, 85,* 465–478.

Juel, C., & Roper-Schneider, D., (1985). The influence of basal readers on first grade reading. *Reading Research Quarterly 20,* 134–152.

Landerl, K., Wimmer, H., & Frith, U. (1997). The impact of orthographic consistency on dyslexia: A German- English comparison. *Cognition, 63*, 315–334.

Makar, B. W. (1995). *Primary phonics.* Cambridge, MA: Educators Publishing Service, Inc.

Maslen, B. L. (1987). *Bob books*. New York: Scholastic.

Mathes, P. G., Howard, J. K., Allen, S. H., & Fuchs, D. (1998). Peer-assisted learning strategies for first-grade readers: Responding to the needs of diverse learners. *Reading Research Quarterly, 33*, 62–94.

Menon, S., & Heibert, E. H. (1999). *Literature anthologies: The task for first-grade readers* (CIERA Report # 1–009). Ann Arbor, MI: Center for the Improvement of Early Reading Achievement.

Messmer, H. (in press). Text decodability and the first-grade reader. *Reading and Writing Quarterly*.

Moustafa, M. (1997, October/November). Bravo to Allington. *Reading Today*, p. 18.

National Reading Panel. (2000). *Summary report*. Bethesda, MD: National Institute of Child Health and Human Development.

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

O'Connor, R. E., Fulmer, D., Harty, K., & Bell, K. (2001, April). *Total awareness: Reducing the severity of reading disability.* Paper presented at American Educational Research Conference, Seattle, WA.

O'Connor, R. E., & Jenkins, J. R. (1999). The prediction of reading disabilities in kindergarten and first grade. *Scientific Studies of Reading, 3*, 159–197.

Paulesu, E., Demonet, J. F., Fazio, F., McCrory, E., Chanoine, V., Brunswick, N., et al. (2001). Dyslexia: Cultural diversity and biological unity. *Science, 291,* 2165–2167.

Pinnell, G. S., Lyons, C. A., DeFord, D. E., Byrk, A. S., & Seltzer, M. (1994). Comparing instructional models for the literacy education of high-risk first graders. *Reading Research Quarterly, 29,* 9–39.

Rayner, K., Foorman, B. R., Perfetti, C. A., Pesetsky, D., & Seidenberg, M. S. (2001). How psychological science informs the teaching of reading. *Psychological Science in the Public Interest, 2*, 31–74.

Rosner, J. (1979). *Helping children overcome learning disabilities (2nd ed.)*. New York: Walker.

Share, D. L. (1995). Phonological recoding and self-teaching: Sine qua non of reading acquisition. *Cognition, 55*, 151–218.

Slavin, R. E., Madden, N. A., Karweit, N. L., Livermon, B. J., & Dolan, L. (1990). Success for all: First-year outcomes of a comprehensive plan for reforming urban education. *American Educational Research Journal, 27*, 255–278.

Snow, C. E., Burns, S., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.

Stahl, S. A., & Huebach, L. (in press). Fluency oriented reading instruction. *Elementary School Journal*.

Stein, M., Johnson, B. & Gutlohn, L. (1999). Analyzing beginning reading programs. *Remedial and Special Education, 20*, 275–287.

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1999). *Test of Word Reading Efficiency*. Austin, TX: PRO-ED.

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

Tunmer, W. E., & Chapman, J. W. (1998). Language prediction skill, phonological recoding ability, and beginning reading. In C. Hulme & R.M. Joshi (Eds.), *Reading and spelling: Development and disorders* (pp. 33–67). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

U.S. Department of Education. (1996). Mapping out the national assessment of Title I: The Interim report [Electronic version]. Retrieved April 16, 2002, from http://www.www.ed.gov/pubs/NatAssess/cover.html

U.S. Department of Education. (2000). Twenty-second annual report to Congress on the Implementation of the Individuals with Disabilities Education Act [Electronic version]. Retrieved April 19, 2002 from http://www.ed.gov/offices/OSERS/OSEP/Products/OSEP2000AnlRrpt/ExecSumm.html

Vadasy, P. F., Jenkins, J. R., Antil, L. R., Wayne, S. K., & O'Connor, R. E. (1997a). Community-based early reading intervention for at-risk first graders. *Learning Disabilities Research & Practice, 12*, 29–39.

Vadasy, P. F., Jenkins, J. R., Antil, L. R., Wayne, S. K., & O'Connor, R. E. (1997b). The effectiveness of one-to-one tutoring by community tutors for at-risk beginning readers. *Learning Disabilities Quarterly, 20,* 126–139.

Vadasy, P. F., Jenkins, J. R., & Pool, K. (2000). Effects of a first-grade tutoring program in phonological and early reading skills. *Journal of Learning Disabilities, 33*, 579–590.

Vadasy, P. F., Wayne, S., O'Connor, R. E., Jenkins, J. R., Pool, K., Firebaugh, M., et al. (2000). *Sound partners: A reading tutoring program for first-grade at-risk readers*. Seattle: Washington Research Institute.

Vellutino, F. R., Scanlon , D. M. , Sipay, E., Small, S., Pratt, A., Chen, R., et al. (1996). Cognitive profiles of difficult-to-remediate and readily remediated poor readers: Early intervention as a vehicle for distinguishing between cognitive and experiential deficits as basic causes of specific reading disability. *Journal of Educational Psychology, 88*, 601–638.

Wagner, R., Torgesen, J. K., & Rashotte, C. A. (1999). *Comprehensive Test of Phonological Processing*. Pro-Ed: Austin, TX.

Wasik, B. A. (1998). Volunteer tutoring programs in reading: A review. *Reading Research Quarterly, 33,* 266–292.

Wimmer, H., & Mayringer, H. (2001). Is the reading-rate problem of German dyslexic children caused by slow visual processes? In M. Wolf (Ed.), *Dyslexia, fluency, and the brain (*pp. 93–102). Timonium, MD: York Press.

Woodcock, R. W. (1987). *Woodcock Reading Mastery Test-Revised*. Allen, TX: DLM.

Wright Group. (1996). *Sunshine*. Bothell, WA: Author.

Wright Group. (1997). *Vision Series.* Bothell, WA: Author.

Wright Group. (1999). *The Wright Skills*. Bothell, WA: Author.

Wright Group. (2000). *The Story Box*. Bothell, WA: Author.

Yopp, H. K. (1988). The validity and reliability of phonemic awareness tests. *Reading Research Quarterly, 23*, 159–177.

Manuscript received November 7, 2002
Accepted June 12, 2003