Journal of Educational Psychology
2006, Vol. 98, No. 4, 690–713

0022-0663/06/$12.00 DOI: 10.1037/0022-0663.98.4.690

Teaching Phonics in the Context of Children's Literature or Spelling: Influences on First-Grade Reading, Spelling, and Writing and Fifth-Grade Comprehension

Theresa A. Roberts and Anne Meiring
California State University, Sacramento

First-grade children's reading, writing, and spelling competencies in 2 different instructional contexts for teaching phonics were examined. Reading, writing, and spelling abilities were measured at the beginning, middle, and end of 1st grade. Children were randomly assigned to 1 of 2 treatments designed to teach grapheme–phoneme correspondences, blending, and segmenting. In 1 treatment, children generated spellings for words, and in the other treatment, phonics instruction was embedded in literature. The spelling treatment was significantly better for spelling phonetically regular real and pseudowords, reading phonetically regular pseudowords, and written story length. It was also more beneficial for low-ability children's reading of connected text. There were no treatment effects on reading uncontrolled words in text. At the end of 5th grade, spelling-context children had significantly higher comprehension than did literature-context children. Discussion focuses on phonological processing while spelling and the effects of the instructional press of context-embedded and context-reduced instructional approaches in 1st grade.

Keywords: spelling instruction, literature-based instruction

There is abundant evidence that phonics instruction is crucial in developing beginning literacy competence (e.g., Adams, 1990; Ball & Blachman, 1988; Bradley & Bryant, 1983; Byrne & Fielding-Barnsley, 1989; Foorman, Francis, Novy, & Liberman, 1991; McGuiness, McGuiness, & Donahue, 1995; NICHHD, 2000; Torgeson et al., 1999; Tunmer, Herriman, & Nesdale, 1988). In their meta-analysis, the National Reading Panel (NICHHD, 2000) concluded that although systematic phonics programs were significantly better than nonphonics programs, there was no evidence of superiority of any one type of phonics program or any one specific program. The report noted 13 important variations in both what is taught and how children are taught phonics. Yet comparisons were only made among three types of programs because of insufficient detail in study reports. The three types of programs that were examined by the National Reading Panel were (a) synthetic phonics (letter–sound correspondences to words), (b) analysis and blending of larger subparts of words (onsets–rimes, spelling patterns), and (c) other programs that did not fit into either of the other two groups. Clearly, many questions remain about the effects on literacy acquisition of variations in classroom instructional context for teaching phonics. Real-world instructional settings are complex, and instructional programs differ on many variables.

One of these variations that is particularly interesting is the degree of embeddedness of the phonics instruction. For the purposes of this study, *embeddedness* refers to the degree to which the phonics instruction occurs in the context of actual text. In a highly embedded context, phonics instruction would be linked to, for example, reading of children's literature or other text material that children are engaging with for the purpose of developing meaning. In a disembedded context, phonic skills would be taught with individual words decontextualized from text selections. Two theoretically plausible contexts for effective teaching of systematic phonics that are also useful for investigating the influence of degree of embeddedness of phonics instruction are the context of reading children's literature and the context of generating word spellings. The focus of the present study was to compare the effects of phonics instruction in these two contexts on reading, spelling, and writing. The rationale for the selection of these two contexts follows.

Literature Context

Interest in teaching reading through the use of children's literature continues to be high in both the research community (Morrow, 1992; Neumann, 1999) and among practitioners (Routman, 1994). To be sure, there is evidence that the use of children's literature confers a number of literacy benefits. Children's reading of literature can increase their use of comprehension strategies and motivation to read (Guthrie, Wigfield, & VonSecker, 2000), understanding of text structure (Morrow, 1992), and vocabulary knowledge (Elley, 1989; Pemberton & Watkins, 1987; Robbins & Ehri, 1994; Senechal & Cornell, 1993; Senechal, Thomas, & Monker, 1995). In addition to evidence from research studies, philosophical orientations that emphasize the meaning-making purpose of reading and the holistic nature of reading, in which many competencies must be orchestrated in unison, have noted the importance of literature in learning to read (e.g., Bergeron, 1990; Goodman, 1986; Weaver, 1994).

Theresa A. Roberts, Department of Child Development, California State University, Sacramento; Anne Meiring, Department of Teacher Education, California State University, Sacramento.

Correspondence concerning this article should be addressed to Theresa A. Roberts, Department of Child Development, California State University, Sacramento, 6000 J Street, Sacramento, CA 95819-2694. E-mail: robertst@csus.edu

Proponents of holistic programs have argued that the phonological skills necessary for skilled reading such as grapheme–phoneme correspondences and blending can be learned within the activities of a print-rich instructional program that utilizes literature, integrates reading, writing, and spelling, and provides for literacy activities that are authentic and meaningful. No direct teaching is necessary or desirable in the strongest form of this persuasion (e.g., Goodman, 1986; Gunderson & Shapiro, 1988). Other writers have suggested that the teaching of phonics may be acceptable but that it should occur in the context of engaging with meaningful literature, with any phonics instruction guided by a child's need to know (Holdaway, 1979; Newman & Church, 1990; Strickland, 1998). It is believed that when word reading instruction occurs in the course of reading connected text, the holistic nature of skilled reading—in which comprehension and word-level decoding skills are orchestrated together for the purpose of obtaining meaning—is preserved. In contrast, when this integrated, holistic process is broken down and subcomponents such as phonics become the object of instruction removed from text, the authentic purpose of decoding is lost and comprehension is disrupted.

More recent treatments of skill instruction and in particular phonics instruction by those emphasizing the holistic perspective have taken the more moderate view frequently described as balanced or comprehensive literacy instruction (International Reading Association and National Association for the Education of Young Children, 1998; Weaver, 1994). Yet graphophonemic cues are still often viewed as a last resort to be used only if semantic and syntactic efforts fail. Recommendations to rely only on partial letter–sound knowledge, on initial or final consonants, or to determine whether the word "looks right" are often made (Adams, 1998). According to these views, teaching beginning phonics in the context of children's literature could be expected to promote the use of phonic knowledge during text reading because decoding is embedded in the context in which it is to be used, thereby promoting generalization of taught knowledge. It also might deepen children's understanding that mastering the alphabetic principle (e.g., that letters map sounds in word pronunciations) helps in accessing the meaning of printed language, thus providing motivation and understanding of the functional and authentic value of learning to decode. Systematic instruction in beginning phonics embedded in literature then could be expected to foster children's ability to read words in connected text and to increase comprehension by advancing understanding of the utility of word decoding skill for comprehension.

Yet there has been limited direct experimental investigation on the effectiveness of teaching phonics in the context of children's literature. Chall (1983) examined studies comparing systematic phonics instruction with those with intrinsic phonics instruction. Intrinsic programs were those in which phonics was presented with less direct emphasis, typically in the context of connected text in which the primary instructional focus was on meaning. She reported advantages on word recognition, oral reading, spelling, vocabulary, and comprehension for the more systematic phonics programs that tended to rely on skills taught separately—decontextualized—from text selections. Foorman, Francis, Fletcher, Schatshneider, and Mehta (1998) reported that a larger portion of children receiving Title 1 services benefited from a direct instruction program in comparison with either a whole-language or embedded phonics program in which phonics skills were taught embedded in predictable books and library books. Torgeson et al. (1999) compared an intensive phonological awareness and synthetic phonics program with several others including one that had systematic but less intensive phonics and more comprehension instruction. They found good evidence of phonological awareness and synthetic phonics superiority on word reading and spelling but less consistent benefit on comprehension. Another study designed to compare code-oriented and meaning-oriented beginning reading instruction in which the treatments were operationalized largely as an embedded or disembedded approach reported advantages for the code-oriented program (Center, Freeman, & Robertson, 2001). A number of methodological limitations render this study inconclusive (e.g., no random assignment or pretests to equate groups and within-school implementation). A very recent study by Christensen and Bowey (2005) also reported that explicit instruction in grapheme–phoneme correspondence was superior to an implicit approach in which phonics was taught embedded in children's literature on measures of reading, spelling, and immediate sentence comprehension. Elements of instruction were matched at a detailed level including amount and sequence of phonics taught and instructional strategy.

Together, these studies reveal that both the quantity and the qualitative nature of phonics instruction influence literacy acquisition. However, in both the Foorman et al. (1998) and Torgeson et al. (1999) studies, the embedded and explicit programs differed on a number of dimensions including amount of phonics and comprehension instruction, amount and type of reading and writing experiences, and types of books that were used. Trained tutors were used for small-group instruction as well, and only selected first-grade children were included: Title I and phonologically disabled, respectively. Stahl, Duffy-Hester, and Stahl (1998) made the point that differences in the amount of phonic content alone may account for outcome differences among phonics programs. For example, synthetic phonics programs may cover more phonic content than other programs. Although Christensen and Bowey (2005) had similar controls to ours for the sequencing of instruction and words included in instruction, their intervention was implemented in small groups (6–8 students) of older (second-grade) children characterized as "advanced beginning readers" (p. 330) taught by trained research assistants. Children in the explicit grapheme–phoneme treatment also received more practice in reading and writing words than did children in the implicit condition. In addition, writing was not evaluated in their study, and there was no longitudinal evaluation of reading comprehension.

In the present study, embedded phonics (literature context) and disembedded phonics instruction (spelling context) were matched on a number of important variables including quantity of letter–sound and blending instruction using a synthetic phonics approach, the amount of segmenting sounds in spoken words, the sequence and number of phonic elements taught, the number of word exemplars of each phonic element introduced in instruction, the requirement of concretely observable responses from children, and utilization of orthographic representations of exemplar words during phonics instruction. In addition to these controls in the phonics instruction, children in both groups participated in the same literature-based program for the nonphonics block of classroom literacy activity. Children were randomly assigned to treatments across classrooms. Regular classroom teachers rather than trained tutors delivered the instruction to all enrolled children in class-

rooms with a 30:1 ratio. Thus, this study was designed to be a more fine-grained implementation and examination of embedded and disembedded phonics instruction than has been previously reported.

Spelling Context

Recent research identifies the reciprocal nature of reading and spelling (Ehri, 1992, 1997; Foorman & Francis, 1994; Gough, Juel, & Griffith, 1992). Part of the theoretical explanation for this phenomenon is that knowledge of the orthography of the language is essential in both decoding words while reading and encoding word representations while spelling (Ehri, 1989, 1997; Richgels, 1995; Tangel & Blachman, 1995; Uhry & Shepherd, 1993). For example, Ehri and Wilce (1980, 1987a) found that word decoding practice improved the spelling of kindergarten and first-grade children.

Other studies attest to the importance of phonological processing in both spelling and writing acquisition (Chomsky, 1979; Clarke, 1988; Ehri, 1989). These studies show that the development of spelling ability can be characterized as moving from letter awareness to partial phonetic and finally to phonetic principles and that at even the earliest stages children are utilizing phonetic cues to guide their developing competencies in writing (Clarke, 1988). Additional evidence reveals that children progress faster in reading and spelling if they receive spelling instruction in the early grades (O'Connor & Jenkins, 1995; Tangel & Blachman, 1995; Uhry & Shepherd, 1993). Instruction in spelling patterns was included in these studies.

Studies on inventive spelling in which children are encouraged to attempt to spell words using their knowledge of letter names and letter sounds have also shown that such experiences contribute to both spelling and reading development (Clarke, 1988). In a review of spelling development of primary-grade children, Graham (2000) concluded that first-grade children do as well or better on spelling tests when using an implicit "natural" spelling approach compared with more structured approaches that specifically teach correct spellings, even though there has been limited research on the issue.

Clark (1988) compared traditional and inventive spelling in middle-class first graders. She reported advantages for the inventive spelling group on spelling regular words, reading regular and irregular words and pseudowords, and written story length. At the end of treatment, there were no differences between groups on comprehension. Similar results were reported by Uhry and Shepherd (1993) in a study in which middle-class children received spelling instruction coupled with the use of inventive spelling in writing. The spelling group had superior word reading, spelling, and segmenting. Instruction occurred in small groups taught by a tutor in which words were segmented and spelled. Word dictation by computer was included. In the comparison group, children read the same words and had other games implemented on the computer. These authors also reported no treatment effects for silent comprehension. Uhry and Shepherd suggested that development of word reading fluency takes time and, consequently, evidence of comprehension benefits from better word reading and spelling afforded by spelling instruction may not be immediately evident. In the present study, we examined comprehension both immediately after instruction, 3 months after instruction, and again 4 years later.

Ehri and Wilce (1987b) and Ehri (1989) also showed how spelling contributes to reading. Ehri and Wilce (1987b) had 4-year-old prereaders either practice spelling words or match letter tiles to sounds. Spelling-trained children performed better on word reading. Ehri and Wilce concluded that spelling helped segmentation, which is important because it shows awareness of speech-to-print relationships. The positive influence of phonemic awareness training on inventive spelling reported by Tangel and Blachman (1995) in a low socioeconomic status group of inner-city kindergarten children provides further evidence of the link between spelling and phonological components of language. Again instruction was carried out by trained tutors in small groups of 4–5 children. These studies provide converging evidence that in small-group settings, spelling instruction consistently contributes to phonetic word spelling and word reading but does not appear to benefit immediate silent reading comprehension. All of these studies utilized small-group instruction with specially trained tutors and thus did not test what might be accomplished by regular teachers with a full class of students.

We were interested in examining whether children's attempts to spell words in lieu of instruction in specific spelling patterns would influence their reading and spelling. Our reasoning, like that of others (Adams, 1990; Stahl & Murray, 1998; Treiman, 1998), was that as children attempt to generate spellings for words (beginning with just initial and final consonants), they must make decisions about relationships between the phonological information they receive in the oral pronunciation of the word and the orthographic (written) form that represents the sounds they have isolated. This decision making fosters an analytic orientation to word structure that speeds acquisition of the alphabetic principle. In addition, children must segment phonemes when they take an oral pronunciation of a word and assign written letters to each sound they hear in the word. Segmentation ability has been shown to be a higher level component of phonemic awareness and closely related to reading ability (Adams, 1990; Yopp, 1988). A very recent study (Silva & Martins, 2003) reported significant advantages on initial phoneme identity and deletion and word segmentation from invented spelling efforts. Thus, it is possible that linking letter–sound correspondence instruction with attempts to spell words that include target letter–sound correspondences will enhance both children's word reading and spelling performance by both teaching them letter–sound correspondences and affording practice in analysis of the phonemic structure of words. Learning phonics through spelling experiences could also be expected to support comprehension through its contribution to automatic word recognition.

The question of the relative efficacy of explicit phonics instruction delivered in the context of children's literature in comparison with explicit phonics instruction in the context of individual words decontexualized from meaningful text remains an open question. And it is a question of significant interest to practitioners and researchers alike. In the present study, we attempted to isolate the embeddedness factor by comparing phonics delivered in the context of children's literature with phonics taught in the context of attempts to partially spell (69 lessons) or completely spell (14 lessons) individual words. The two groups were equated on the sequence of grapheme–phoneme instruction, amount of blending and segmenting, number of direct instruction exposures to the target letter–sound correspondences, observable concrete responses from children, and active utilization of phonics knowledge

in written whole words. What differed was whether phonics instruction was embedded in children's literature or disembedded from text in efforts to spell individual words. All children in two regular first-grade classrooms were included as participants, and instruction was delivered in a whole-class setting with a 30:1 student–teacher ratio. In addition, regular classroom teachers delivered the instruction. Each teacher taught one half of the lessons in each treatment. Thus, this study that examined the context of phonics instruction more closely matched the conditions of typical classroom instruction than previous studies. Including children of all ability levels in the study rather than only children at risk, as has been done in most previous studies, allowed us to also examine the effects of context variation on low-, average-, and high-ability children. Spelling, reading, writing, and both contemporaneous comprehension and comprehension 4 years later were also examined in one study.

Method

Participants

Sixty-one children enrolled in two first-grade classrooms were initially included in the study. Five children moved during the course of the year (3 from the spelling treatment, 2 from the literature treatment), and 1 child was eliminated for excessive absences (spelling group). Forty percent of the participating children were African American, 30% were Filipino, 11% were White, 9% were Hispanic, 5% were Asian, and 5% were classified as "other." Forty-eight percent of the children were from families receiving welfare or free lunches. Eighteen percent of the students were identified as English language learners. Nonparametric tests showed that there were no differences between treatments in terms of ethnicity and English-language-learner status. Children, blocked by high, medium, and low initial literacy competence based on letter name and letter sound pretests given at the beginning of first grade, were randomly assigned to one of two groups that were then assigned to one of the two treatments. Each child was individually administered four reading tasks, group administered three spelling tasks, and a writing task at the beginning, middle, and end of first grade.

Measures

Measures sensitive to possible benefits of both treatments were selected. Spelling and reading tasks requiring performance on isolated words (pseudoword reading and spelling, real and sight word spelling, familiar story words on word cards) and words in meaningful contexts (reading words in familiar and unfamiliar stories, correct spelling in student stories, writing words in stories) were included. The spelling treatment focused on learning phonics through experiences with individual words, whereas the literature treatment relied heavily on learning phonics skills in connected text. We sought to ensure that neither group would be favored in our assessments by using tasks that were similar to the emphasis of only one treatment.

Spelling tasks. Children were group administered three spelling tests of 20 words each on each of the three testing occasions during the year: phonetically regular real words, phonetically regular pseudowords, and high-frequency sight words. Three separate real word and related pseudoword lists were constructed and were counterbalanced across testing sessions to control for list and repeated reading effects. Spelling tests were group administered, with children's desks individually separated with small cardboard carrels placed on them to prevent copying. Of the words included on spelling tests, 17 appeared in the spelling-context instruction one time, and 9 were story words used in the literature-context instruction.

Phonetically regular real words. This test included 20 phonetically regular real words using consonant–vowel–consonant (15 words) and consonant–consonant–vowel–consonant (5 words) word patterns using the short vowels /a/ (six words), /i/ (seven words), and /o/ (seven words); (test–retest reliability = .70; see Appendix A).

Phonetically regular pseudowords. These words were formed by transposing the initial and final letter of words appearing on the test of phonetically regular real words, ensuring that both real words and pseudowords had the same letters and visual elements (test–retest reliability = .68). For words with blends, simple transposing resulted in an illegal English spelling so for the consonant–consonant–vowel–consonant words, the letters were arranged to spell a word that did not violate English spelling and pronunciation rules. The pseudoword spelling test was included to obtain a measure of children's ability to use phonics in spelling in which there was no possibility that specific spellings were learned through previous encounters.

High-frequency sight words. A test of 20 sight words drawn from grade-level lists that came with the basal texts used in the children's classrooms was also developed (test–retest reliability = .90). Children were given this same test on each testing occasion, as there were not enough words to create three parallel lists of sight words.

Reading tasks. Three reading tests were administered individually in a set order graduated in familiarity: reading a familiar story, reading familiar story words out of context, and reading the next story from the reading basal not yet read in class. A pseudoword reading task, matched to the spelling tests, was administered separately.

Word reading in context of familiar story. Children individually read a story that they had just completed reading in class 3–5 times with the whole group as part of the regular instructional program. Children's ability to read 26 target words (the total number of different words in the first story) was scored. These words were not controlled for phonetic spellings. The number of target words correctly pronounced at their first appearance in text was counted. This was a measure of children's ability to read words in the context of a familiar story. The first story had a very repetitive structure, and almost all of the word variability was substituting one noun or descriptor with another. Construct validity was examined by correlating the February and May test scores ($r = .55$). By conventional standards, this is a relatively low reliability that undoubtedly reflects the substantial word variability between stories in which no effort was made to equate features of the stories because these texts were selected for their ecological validity and to represent reading in context from authentic classroom materials. Given that two different stories were used and that testing occasions were separated by 3 months, this correlation reveals relatively consistent performance. Selection of words from the February and May text selections was guided by the nature of the words from the first (October) story. Of the 26 words selected for each story, 13–19 were nouns and verbs and 7–13 were descriptors–modifiers (see Appendix B). Miscues were recorded as children read in order to examine the types of word reading errors children made and word reading strategies they used. Mispronunciations, no attempts, and hesitations were recorded.

Decontextualized word reading. Children read the 26 target words from the October, February, and May familiar story presented in lowercase letters on 3 in. × 5 in. cards (76.2 mm × 127 mm) one at a time (test–retest reliability = .89) as a measure of decontextualized word reading. This task served as a measure of context-free reading of the same words that had already been read in context with the teacher as part of the regular instructional program.

Word reading in context of unfamiliar story. A third task, similar to the first, required children to orally read the next selection from their reading text that was to be read as a whole group and individually in class but which had not yet been read (test–retest reliability = .87). This task was a measure of independent reading in context. The Time 2 and Time 3 words, respectively, were similar types: 8 and 10 verbs, 11 and 8 nouns, and 7 and 8 descriptor–modifiers (see Appendix A). Text selections were 167 words (the number of words in the first story) to 176 words in length. February

and May selections were a portion of complete stories. The same miscue scoring as described for the familiar story reading was also implemented.

Pseudoword reading. Children read a series of 20 phonetically regular pseudowords composed of consonant–vowel–consonant and consonant–consonant–vowel–consonant word patterns presented in lowercase letters printed on 3 in. × 5 in. cards (76.2 mm × 127 mm; test–retest reliability = .86). These words were the same as the words used for the pseudoword spelling tasks. Three separate lists were constructed and were counterbalanced across testing sessions to control for list and repeated reading effects. This task was conceptualized as a measure of children's ability to read words using a three-part decoding model: parsing letters from written words, converting those letters to sounds, and then blending those sounds into a pronunciation. Pseudoword reading is unlikely to be influenced by previous visual exposure and is considered to be a good measure of decoding ability (e.g., Harris & Sipay, 1990; Perfetti & Hogaboam, 1975; Tunmer & Nesdale, 1985).

Writing tasks. Children were asked to write about one of three different topics during the October, February, and May testing. The topics were toys, recess, and friends. Topics were selected to ensure that all children would have related prior knowledge and that they were relevant to experiences that children were interested in. Children were encouraged to generate spellings for words on their own and were allowed to write until they chose to stop. Topics were counterbalanced across testing occasions. Writing was done as a whole class, with children assigned to write on the appropriate topic. The same individual cardboard carrels as used during spelling tests were used. These stories were scored for number of words as a measure of writing fluency, number of correctly spelled words, and number of idea units represented in the stories. An overall interrater reliability of .97 for scoring the number of words and correctly spelled words was obtained on a randomly drawn sample of 25% of the stories from the middle and end of first-grade test periods. Interrater reliability for number of ideas units was noticeably lower at $r = .61$; therefore, only the fluency scores were used in analyses.

Comprehension tests. The Metropolitan Achievement Test (MAT) in reading comprehension was given at the end of first grade. The Curriculum Aligned Assessment (CAS) standardized test adopted by the district was administered at the end of fifth grade, which was the last year that students attended the elementary school before moving on to a Grade 6–8 middle school. This test included 22 items that assessed literal and inferential comprehension. Literal comprehension items included finding details, identifying main ideas given in text, and answering factual questions. Inferential questions including making interpretations, drawing unstated conclusions, and analyzing character's actions. Both narrative and expository text was included in the assessment.

Spelling and Literature Instruction Procedures

Students were randomly assigned across classrooms to either the spelling-context or the literature-context group. Teachers switched between groups every 4 weeks. Thus, potential classroom and teacher effects were controlled. The second testing occurred after the 83 lessons at the end of February, with retention testing occurring in May. Instruction was designed in light of a three-part model of decoding that recognized the necessary (a) segmenting of individual graphemes in reading and segmenting of individual phonemes in spelling, (b) conversion from grapheme to phoneme in word decoding and conversion from phoneme to grapheme in spelling, and (c) blending phonemes to obtain a word pronunciation for lexical access during reading (literature group) and to check a match between a written word and a stored word pronunciation when spelling (spelling group).

Instruction in the two groups was equated for (a) the length of the lessons (15–20 min), (b) the sequence of letter–sound correspondences taught (*m*, *s*, *f*, *b*, *t*, *c*, *a*, *r*, *l*, *p*, *o*, *d*, *g*, *n*, *w*, *i*, *h*, *j*, *k*, *v*, *u, y*, *e*), (c) the number of phonetically regular exemplars of each target sound that were presented (16), (d) the sequence of regular and review lessons, (e) the requirement of a response by the students to indicate initial or final position of target sounds in exemplar words, (f) using letter–sound correspondences to blend whole words, (g) activity with graphemes represented in whole words, and (h) whole-group instruction given by regular classroom teachers. Table 1 affords a comparison of the instructional steps taken in each treatment. Differences between treatments are shown in boldface.

Forty-four of 83 lessons focused on letter–sound correspondences with either dictation of eight words for writing initial or final sounds (spelling-context group) or reading an excerpt from a storybook that had been read interactively and that contained eight words, with the target sound in initial or final position (literature-context group). Children segmented initial or final sounds in target words by partial spellings (spelling-context group) or by showing one of two cards during a second reading of the selection (literature-context group). Each grapheme–phoneme correspondence was practiced in this way for 2 days for a total of 16 words (44 lessons). A review lesson (25 lessons) occurred after each new letter–sound pair was practiced for 2 days. After introduction of the first six letters, whole words were dictated (spelling-context group) or responded to in a literature selection (literature-context group; 14 lessons). Blending was taught in these whole-word lessons in both groups. The teacher wrote the letters of target words on the board in the whole-word lessons and simultaneously pronounced their sounds followed by choral blending. The same letter–sound correspondence, review lesson, or whole-group lesson was taught each day in the two groups. Practice words used in each type of lesson came from either the lists available in the *Spelling Through Phonics* manual (McCracken & McCracken, 1982) for the spelling group or from the literature selection read to the class for the literature group (see Appendix C).

Students in the spelling group participated in 15- to 20-min daily lessons using words from the Spelling Through Phonics program (McCracken & McCracken, 1982). This program is designed to teach letter–sound correspondences and to develop competence in analysis of the sequence of sounds in words. Conventional spellings patterns are not taught, and words are not organized into any sequence of high-utility spellings. There were many multisyllabic, uncommon, and irregular words to be used for instruction in letter–sound correspondences that we reasoned would be a good match for the range of words to be encountered in the children's literature treatment. The phonics in literature group participated in 15- to 20-min daily lessons in which letter–sound correspondences were taught within the context of children's literature. Picture books selected from the California State Department of Education (1996) document, *Recommended Literature: Kindergarten through Grade Twelve*, were used to teach letter–sound correspondences and blending skills, using the same sequence of letter–sound correspondences that were taught in the spelling group (see Appendix C for a list of the selected books). These literature selections were read with discussion on meaning prior to the phonics instruction. The words for the phonics instruction were presented in an excerpt taken verbatim from the story, and thus the semantic and narrative connection of these words to meaning in an authentic text was ensured.

The rest of the literacy development program in both classrooms was a literature-based program that heavily utilized repeated and choral reading of literature, pocket charts, sentence frames, recitations and singing, book writing, basal texts, and big books. There was daily writing in literature response journals and daily expressive writing. Inventive–transitional spelling was encouraged. Extensive use of oral discussion, sustained silent reading, text-related art projects, and integration of language arts with other curricular areas was also included in both classrooms. The two teachers jointly developed written lesson plans for the nonexperimental portion of the literacy program and devoted a 2-hr block to literacy activity every day. Thus, the two classrooms can be characterized as providing a print-rich environment with a great deal of meaningful literacy activity balanced with systematic, direct instruction in the alphabetic principle for approximately 12%–17% of the total 2-hr literacy block.

Table 1
Lesson Sequence for Each Treatment

Spelling context	Literature context
Individual letter lessons (44 lessons)	
	Interactively reads literature. **Displays story excerpt.**
1. Introduces letter.	1. Introduces letter.
2. "Letter stands for." Children repeat sound. **Teacher draws lines on chalkboard.** **"Set up your chalkboards like this."**	2. "Letter stands for." Children repeat sound. **Teacher introduces story excerpt.** **"While I am reading, listen for /sound/."**
	3. **Teacher reads excerpt.** **"You read with me."** **"Listen for /sound/."**
	4. **Teacher reads excerpt slowly.** **"Raise your hand when you hear /sound/."** **Teacher underlines target letters in word in excerpt.** **Teacher reads excerpt again with students chorally.**
5. "Listen for /sound/." "If hear /sound/ at beginning, write on first line." "If hear /sound/ at end, write on second line."	5. "Listen for /sound/." "If hear /sound/ at beginning, **show your blue card**." "If hear /sound/ at end, **show your yellow card**."
6. Teacher says 4 words one at a time. Children repeat word. **Children write partial spellings for initial or final position.** **Teacher models correct partial spelling.** **Children pronounce word as teacher points to spelling.** **Erase boards.**	6. Teacher says 4 words from the excerpt one at a time. Children repeat word. **Children hold up cards for initial or final position.** **Teacher shows word on card with correct letter underlined.** **Children pronounce word as teacher points to word written on card.**
7. Teacher says 4 words one at a time. **Children write partial spellings.** **Children pronounce word as teacher points to spelling.** **Teacher repeats each word and shows correct partial spelling.**	7. Teacher says 4 words from the excerpt one at a time. **Children show cards for initial or final position.** **Children pronounce word as teacher shows word on card.** **Teacher repeats each word and shows correct letter position on written word.**
8. Children pronounce words as teacher points to spelling.	8. Children pronounce words as teacher points to spelling.
Review lessons (25 lessons)	
	Reads literature. **Shows excerpts.**
1. "Which letters so far?"	1. "Which letters so far?"
2. Teacher writes letters.	2. Teacher writes letters.
3. Review sounds.	3. Review sounds. **"Listen for sounds" in excerpt.** **Teacher reads excerpt.**
4. Teacher says review words one at a time (8 words). **Children write letters for beginning or ending sound.** **Teacher models correct partial spelling.**	4. Teacher says review words one at a time (8 words). **Children show card for beginning or ending sound.** **Teacher models correct segmenting on word card.**
Whole-word lessons (14 lessons)	
1. "Write whole words." **Teacher dictates words one at a time.** **Children repeat (8 words).** **Children write spellings for each word on board.** **Children read spellings.**	1. "Write whole words." **Teacher reads excerpt with target words underlined.** **Teacher points to words in excerpt one at a time (8 words), stretches pronunciation, and blends word.** **Children chorally blend.** **Children read words from story.**
2. Teacher writes word and stretches pronunciation.	2. Teacher writes word and stretches pronunciation.
3. Teacher blends and runs hand underneath with children.	3. Teacher blends and runs hand underneath with children. **Teacher and children read excerpt chorally.**

Note. Differences between treatments are shown in boldface.

Teacher training prior to implementation, scripted lesson plans, and audiotaping of lessons from each treatment were procedures used to ensure fidelity of implementation. Checklists were created by listing the sequential steps taken in the typical lesson from each treatment (these lesson components are shown in Table 1). Lesson audiotapes were listened to with a checklist of the 28 lesson components for the literature lessons and the 19 lesson components for the spelling-context lessons (the difference in number of lesson elements solely was due to additional procedures for handling the story excerpt; see Table 1). A total of 20 lessons was coded in this manner. In the spelling context, 90% of specific lesson elements were

verified, and in the literature context, 91% of lesson elements were verified. In each treatment, one half of the omissions were due to teachers not implementing the final element of lessons, which called for a final pronunciation of each target word by children. Time constraints influenced this omission. Given the detailed level of fidelity analysis performed, these percentages reflect a high level of implementation accuracy. There was 100% agreement between two independent raters on the coding of all of the components of the lessons except whether the story had been read interactively prior to the phonics instruction in the literature-context group and whether children's oral responses could be heard on the tapes in both groups.

Results

There were no significant pretreatment differences between the groups on letter name or letter–sound knowledge, ability to read words in context or presented on cards, or to read words from the next story to be encountered in the program reader (see Table 2). Because there was some missing data across the year, a missing data analysis was performed. The Little McCar test showed that the missing data could be assumed to be randomly distributed, $\chi^2(63) = 57.40$, $p = .67$. Statistical tests revealed that children in both treatment groups showed significant improvement on all of the reading, spelling, and writing tasks from October to February and from February to May, with the exception of reading words in the story that had already been read with the teacher several times. This effect was due to the much higher performance of children on the initial October story that had a very predictable, repetitive structure with only 26 different words. Thus, across the first-grade year, children in both instructional groups were showing signifi-

Table 2
Means, Standard Deviations, and Effect Sizes for Spelling, Reading, Writing, and Comprehension Tasks by Phonics Instruction Context

Measure	Spelling group		Literature group		Effect size
	M	*SD*	*M*	*SD*	
Spelling phonetically regular pseudowords					
October	2.89	5.35	1.93	3.08	
February	11.19	6.50*	4.86	4.86	1.11
May	12.50	5.49*	8.10	5.42	0.81
Spelling phonetically regular real words					
October	3.48	5.77	2.41	3.45	
February	12.04	6.32*	5.79	5.05	1.10
May	12.42	5.39*	9.59	5.88	0.51
Spelling sight words					
October	1.42	2.50	1.38	2.47	
February	4.11	4.13	3.41	4.04	
May	5.72	4.31	5.28	5.06	
Reading of phonetically regular pseudowords					
October	1.85	3.69	1.31	2.52	
February	11.37	6.41*	6.97	6.59	0.68
May	13.38	5.94*	9.29	7.37	0.64
Reading words in familiar story					
October	24.89	1.42	24.07	2.38	
February	21.88	5.43	19.97	6.11	
May	23.04	3.63	22.45	4.28	
Reading words out of context					
October	8.93	8.21	7.87	8.94	
February	12.88	9.15	9.63	9.61	
May	16.12	8.93	14.48	9.32	
Reading words unfamiliar story					
October	9.70	7.68	8.96	7.82	
February	15.42	7.53	12.97	8.21	
May	16.69	8.53	14.97	8.99	
Writing fluency scores					
October	8.34	10.83	6.85	9.68	
February	40.22	30.90	31.21	26.63	
May	45.46	22.13*	35.48	22.41	0.45
Letter names					
October	21.93	5.05	20.93	5.66	
February	25.58	0.70	24.79	2.43	
Letter sounds					
October	12.15	7.04	11.66	7.62	
February	21.92	2.71	20.07	3.49	
MAT comprehension					
February	15.00	7.30	14.58	8.00	
May	17.24	8.41	18.81	8.49	

Note. Effect size was calculated in the following manner: (mean treatment group − mean control group)/pooled standard deviation. MAT = Metropolitan Achievement Test.
* $p \leq .05$.

cant growth in literacy, even though it is not possible to clearly determine whether that growth was a function of instructional experience. Results for children's growth in letter names and sounds, the three spelling tasks, the four reading tasks, the two measures of writing, and standardized comprehension at the end of first grade and fifth grade are discussed in order.

Letter Names and Letter Sounds

A multivariate, repeated measures analysis, with letter names and letter sounds as dependent variables, treatment as a between-subjects factor, and October and February scores as repeated measures, was computed. As a result of the ceiling effect observable on both letter names ($M = 25.02$, $SD = 2.19$) and letter sounds ($M = 20.75$, $SD = 3.64$) in February, letter names and letter sounds were not tested at Time 3. There were no significant differences between the two treatment groups prior to the instructional intervention or in February of the first-grade year. There was only a significant effect for time, with children overall scoring significantly higher in February on both measures (Hotelling's $T^2 = 1.98$), approximate $F(2, 54) = 53.39$, $p < .001$ (see Table 2). Although children began first grade knowing the names of most letters ($M = 21.40$, $SD = 5.34$), their letter–sound scores almost doubled from October to February (October: $M = 12.12$, $SD = 7.26$; February: $M = 20.75$, $SD = 3.64$).

General Approach to Analyzing Performance on Spelling, Reading, and Writing Tasks

Hierarchical longitudinal analyses were used to examine how the literature-context group performed in comparison with the spelling-context group across October, February, and May of first grade on the spelling, reading, and writing tests. Hierarchical linear model (HLM) approaches for analyzing data based on a two-stage model of change that included within-subject growth trajectories and between-subjects correlates of this change were used. Individual and group growth curves can be estimated to describe patterns of change across time and correlates to those changes related to type of instructional experience. HLM was selected for a number of reasons. These models can be used to estimate mean rate of change from one testing phase to another when there are at least three waves of data. HLM is also well suited for data with a nested design and handles missing data well. In the present study, data were collected three times during the first-grade year, children were nested within treatment, and there were some missing data in a small sample size.

Separate HLM analyses using the SPSS mixed model statistical software (Version 13.5) were conducted for each of the measures of (a) sight word, real word, and pseudoword spelling; (b) contextualized and decontextualized reading based on classroom texts and word reading of phonetically regular pseudowords; and (c) story writing. Growth trajectories over the course of first grade were estimated, and the relationship of these patterns of growth to variation in instructional experience was also estimated. The intercept parameter was specified to represent the average ability level of individuals at the onset of data collection. Treatment and time and the Treatment × Time interaction were treated as fixed effects, with time specified as a repeated measure because there were three waves of data collected. Participants were initially treated as a random effect allowing (a) both intercepts and slopes to vary, (b) intercepts only to vary, and (c) slopes only to vary across children. These models examined the degree to which there were important individual differences in initial status (intercept), rates of growth across time (slopes), or the covariance of intercepts and slopes (intercept plus slope). Models were fitted using a variety of options for modeling the within-subject covariance structure. These various models were estimated, and the final model was selected on the basis of minimizing the restricted maximum likelihood estimate (REML) and maximizing Akaike's information criterion (AIC). AIC penalizes models for the number of estimated parameters and thus assists in selecting the simplest model that fits the data well. A particular interest was in determining whether the covariance structures for individuals reflected an autoregressive pattern. An autoregressive covariance structure was initially specified, and if a reasonably good fit for the data was found with this covariance structure, it was retained because it accounts for within-subject correlation of scores at each testing point, a structure well suited for longitudinal data. Table 3 shows the results of the HLM analyses. Data reported for the individual curves represent the random effects, and data shown in the group estimates show the fixed effects. Data are only reported for those parameters included in the final model. Figure 1 shows the growth slopes by treatment for each of the spelling and writing measures. Figure 2 shows the growth slopes for the reading measures. Table 2 shows the means and standard deviations for the letter names and letter sounds, reading, spelling, and writing measures by treatment.

Spelling

HLMs were fitted to the data in accordance with the previous description for the three first-grade testing waves, with individuals nested within treatment for each of the pseudoword, real word, and sight word spelling tests.

Real word spelling. For spelling of consonant–vowel–consonant and consonant–consonant–vowel–consonant real words, the model with an autoregressive covariance structure, with a repeated measure for time (October, February, May) and in which both intercepts and slopes for individuals were allowed to vary, fit the data best. There were significant fixed effects for the intercept, treatment, time, and the Treatment × Time interaction. The group curve section of Table 3 shows these results. The significant intercept shows that the average score at initial testing was significantly different from zero. The overall time effect was significant showing that the overall growth slope was significantly greater than zero. Average rate of growth between Time 1 and Time 2 was 7.17 words spelled correctly and between Time 2 and Time 3 was 3.38 words spelled correctly. The significant Time × Treatment interaction was due to the significantly better performance of spelling-context children at Time 2 but not at Time 1 or Time 3. There were significant differences in rate of growth between children in the two treatment groups in favor of the spelling-context group from October to February but not from February to May. On average, spelling-group children improved 7.11 more words than literature-group children from October to February and improved 1.93 words more from February to May. After instruction ended at Time 2, growth in real word spelling for the spelling group slowed even though it was still in advance of the scores of the literature-context group.

(*text continues on page 701*)

Table 3
Hierarchical Linear Model Analyses: Estimates of Individual and Group (Treatment) Growth Curves From October to May of First Grade

	Spelling measures			Reading measures				Writing measure
Measure	Real phonetic	Pseudoword phonetic	Sight	Isolated word cards	Familiar story	Unfamiliar story	Pseudowords phonetic	Fluency
Individual curve								
Intercept			3.08 (1.66)†	67.82 (13.86)***		50.61 (10.75)***		
Slope								
Intercept + slope	9.97 (3.94)*	7.58 (3.49)*			1.22 (0.52)*		4.50 (0.91)***	
ARH1 rho	.61**	.46*	.88***	.86***	.58***		.93***	
Group curve								
Intercept	2.41 (0.99)*	1.93 (0.97)*	1.38 (0.46)***	7.87 (1.66)***	24.07 (0.75)***	6.80 (1.51)***	1.31 (1.03)*	6.85 (1.98)***
Treatment	1.07** (1.42)	0.96 (1.33)**	0.03 (0.66)	1.06 (2.36)	0.83 (0.51)	−2.32 (0.64)	0.54 (0.84)***	1.53 (2.82)*
Testing time								
Initial to Time 2	7.17 (0.97)***	6.17 (0.90)**	2.03 (0.89)**	6.42 (1.10)***	−4.10 (0.85)**	6.17 (1.00)***	5.65 (1.07)***	28.63 (4.58)***
Time 2 to Time 3	3.38 (0.93)***	2.93 (1.01)***	1.86 (0.61)***	1.77 (0.90)†	−1.56 (1.00)	1.97 (1.34)***	2.37 (1.13)***	24.35 (5.69)***
Time × Treatment								
Time 1 × Treatment	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*
Time 2 × Treatment	5.18 (1.28)***	5.36 (1.45)***	0.67 (0.61)	1.42 (1.32)	1.03 (1.24)	−0.87 (1.45)	3.86 (1.54)*	7.48 (8.19)
Time 3 × Treatment	1.93 (1.60)	3.66 (1.30)*	0.80 (0.88)	1.06 (2.36)	−0.24 (1.45)	−0.89 (1.93)	3.68 (1.60)*	8.45 (6.65)
REML	967.85	969.02	721.92	1,067.71	856.75	1,041.18	919.40	1,397.51
AIC	977.85	979.02	731.92	1,075.71	866.75	1,051.18	929.54	1,403.51

Note. Values in parentheses are standard errors. The statistics under *Individual curve* describe individual differences in initial status (*Intercept*), rates of rate of growth (*Slope*), or both initial status and rate of growth (*Intercept + slope*). The major findings of treatment effects across time are indicated under *Group curve,* where *group* refers to treatment differences. The Time × Treatment statistics are those that show the pattern of treatment effects from Time 1 (October) to Time 2 (February) and from Time 2 (February) to Time 3 (May). ARH1 = heterogeneous first-order autoregressive; REML = restricted maximum likelihood estimate; AIC = Akaike's information criterion.
$^{\dagger}p \leq .10$. $^{*}p \leq .05$. $^{**}p \leq .01$. $^{***}p \leq .001$.

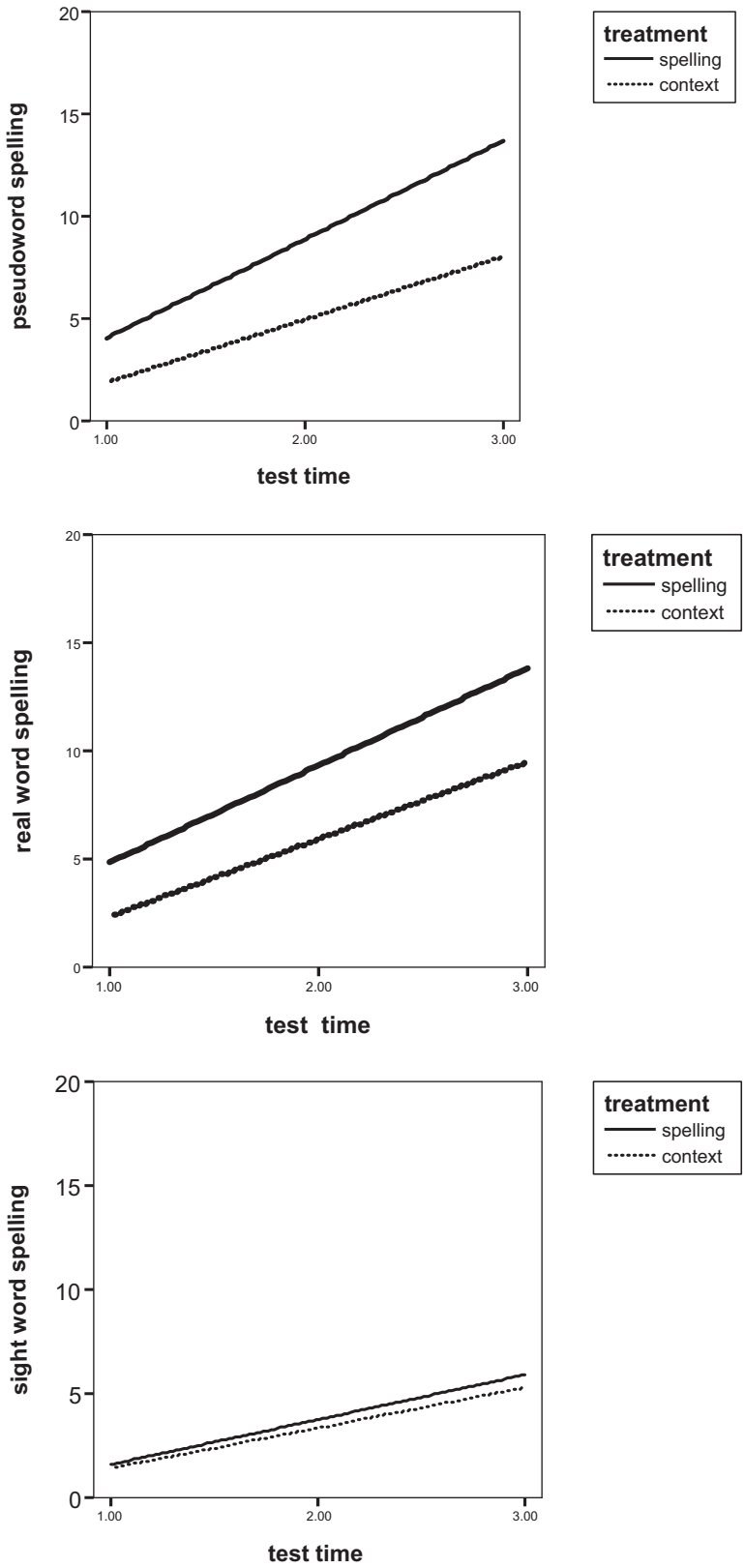

Figure 1. Growth curves by treatment for the spelling measures. Mean scores appear on the *y*-axis. For test time, 1.00 = October of first grade, 2.00 = February of first grade, and 3.00 = May of first grade.

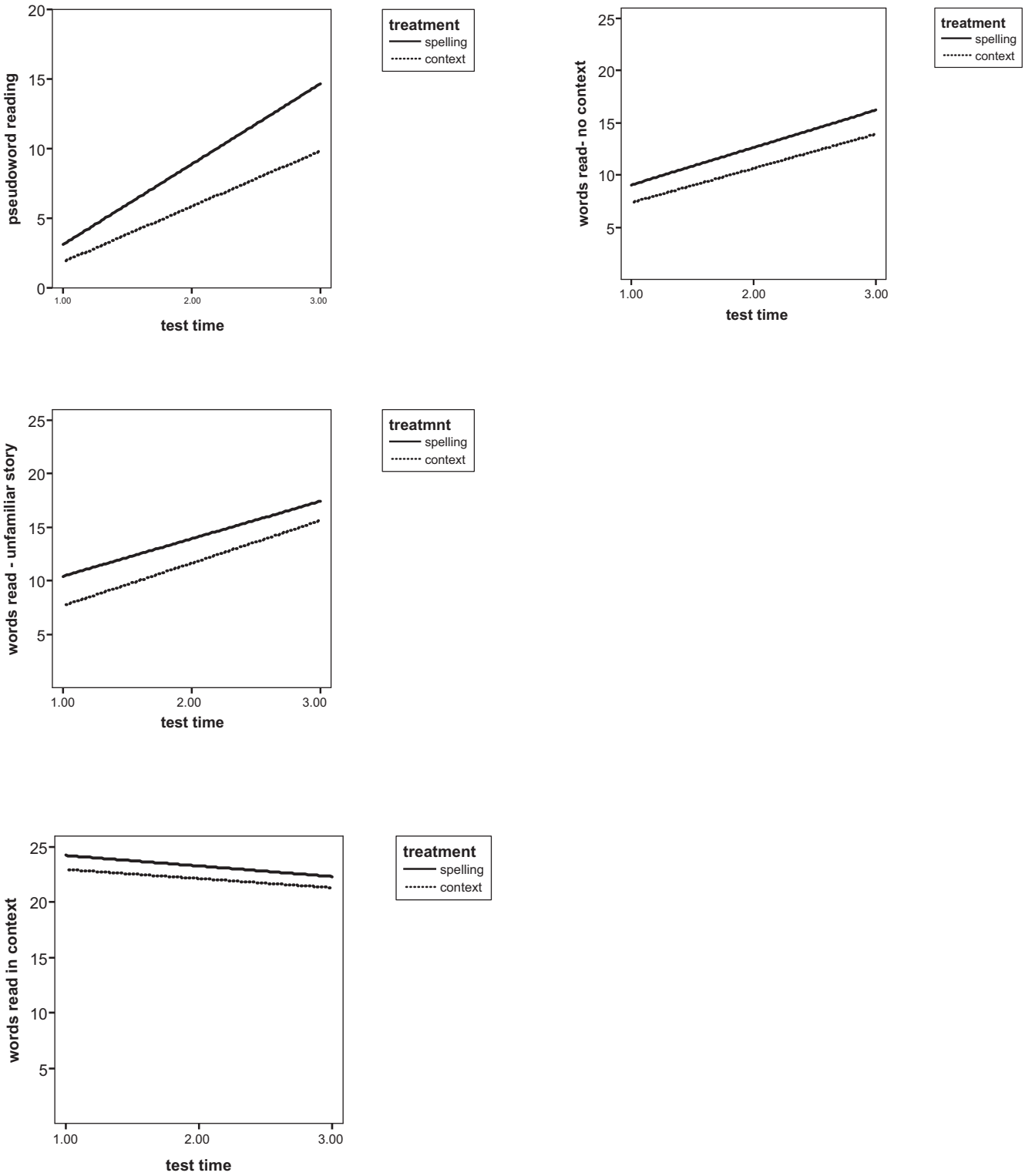

Figure 2. Growth curves by treatment for the reading measures. Mean scores appear on the *y*-axis. For test time, 1.00 = October of first grade, 2.00 = February of first grade, and 3.00 = May of first grade.

The effect size ([control mean – experimental mean]/pooled standard deviation) for Time 2 (February) was 1.10 and was 0.51 for Time 3 (May). The estimate of the random effect for covariance of individual differences in initial status and individual differences in growth trajectories fit the data well (see Table 3). Children who began with higher real word spelling had higher rates of growth (Wald $Z = 2.53$, $p = .02$). The heterogeneous autoregressive effect was significant, showing an overall correlation of .61 between individual variances at the three times of data collection. Children's initial real word spelling was significantly correlated with subsequent spelling scores. The individual curve section of Table 3 presents these results. Figure 1 shows the estimated growth trajectories for the spelling, reading, and writing measures for each treatment.

Pseudoword spelling. On the basis of selection criteria of the smallest size of REML and maximizing AIC, the model that specified a heterogeneous autoregressive covariance structure, with time as a repeated measure and slopes and intercepts for individuals allowed to vary as a random effect, resulted in the best fit to the data. The fixed effects for the intercept, treatment, test time, and the Treatment × Test Time interaction were significant (see Table 3). The significant effect for the intercept shows that children's average initial scores were significantly different from zero. The overall time effect was significant, indicating that the growth slope for pseudoword spelling was overall significantly greater than zero. Between Time 1 and Time 2, children made a mean growth of 6.17 pseudowords spelled correctly, and between Time 2 and Time 3, children made an average growth rate of 2.93 pseudowords. This pattern of deceleration in rate of growth occurred after the end of treatment in February. The Time × Treatment interaction is explained by the similar performance of children at Time 1, whereas there were significant differences in growth slopes in favor of the spelling group at Time 2 and Time 3. The effect sizes ([control mean – experimental mean]/pooled standard deviation) estimated from marginal means shown in Table 2 were approximately 1.11 for Time 2 and 0.81 for Time 3.

The random effect for individuals, with slopes and intercepts varying, was also significant (Wald $Z = 2.22$, $p = .03$), indicating that there was significant variability in individual growth curves, with initial levels of pseudoword spelling correlated with growth across time. In other words, children who started off higher showed greater growth than other children. The heterogeneous first-order autoregressive rho correlation is an index of correlation of variances across time in which the variances from one time to another are not assumed to be equal. The estimate for the correlation of individual variance between testing times was .46 and was statistically significant.

Sight word spelling. A model, with time of testing as a repeated measure, with an autoregressive covariance structure and that also specified the intercept as a random effect minimized REML and maximized AIC on this measure. For fixed effects, there was a significant effect for the intercept and sight word spelling scores across the testing times. The intercept fixed effect result indicates that the average sight word spelling score at Time 1 was significantly different from zero. The mean sight word spelling score at Time 1 was 1.38 ($SD = 0.46$) words. These results can be seen in Table 3 in the section of the table showing estimates for group curves. From Time 1 to Time 2, children on average increased 2.03 correctly spelled sight words, and from Time 2 to Time 3, they increased an average of 1.86 correctly spelled sight words (see Table 3). This increase was significantly different from zero at both testing times. There was no significant effect for treatment or the Treatment × Ability interaction (see Table 3). Table 2 shows that the means for the two treatments on sight word spelling were nearly identical in October, February, and May.

Including the random effect for individual differences in intercepts in the model improved the fit of the model even though the test for the significance of this parameter was only marginally significant (Wald $Z = 1.86$, $p = .06$). The repeated effect for time showed a highly significant heterogeneous autoregressive covariance correlation of .89 for the three testing waves (Wald $Z = 29.01$, $p < .000$), showing substantial correlation between individual scores across time.

Word Reading

Individual HLM analyses analogous to those for the spelling measures were fitted to the data for the three first-grade testing waves, with individuals nested within treatment. One set of HLM analyses was performed for each of the four reading tests. Results for each of the four reading measures follows and can be seen in Table 3 and Figure 1.

Pseudoword reading. Models with fixed effects for the intercept, treatment, time, and the Treatment × Time interaction were specified with random effects estimated for individual variations in intercept, slope, and intercept plus slope in different models. Time was treated as a repeated measure. The model with a random effect for the intercept and time of testing, with a heterogeneous autoregressive covariance matrix, fit the data best. The mean initial performance level for children shown by the intercept parameter estimate (see Table 3) was very low (1.31), even though this mean score was significantly different from zero. There was also a significant main effect for time, showing that growth slopes were significantly greater than zero from Time 1 to Time 2 and again from Time 2 to Time 3. Children had a mean growth rate of 5.65 pseudowords read correctly from October to February and an average growth rate of 2.37 words read correctly from February to May. This pattern shows again a deceleration of growth after instruction stopped in February. There was no significant overall effect of treatment, but the parameter estimate for the Treatment × Time interaction was significant. As with the phonetic spelling measures, there were no differences at Time 1, but at Time 2 and Time 3, children in the spelling-context group had higher rates of growth than children in the literature-context group. The effect size for reading phonetically regular pseudowords ([control mean – experimental mean]/pooled standard deviation) in favor of the spelling treatment estimated from the means shown in Table 2 was .68 at Time 2 and was .64 at Time 3.

The random effect for the intercept plus time was significant (Wald $Z = 4.96$, $p < .001$), showing that children higher in initial status showed greater growth rates in pseudoword reading (see Table 3). The covariance of scores within individuals across testing times was highly significant, as indicated by the heterogeneous autoregressive correlation (ARHI rho; see Table 3) of .93. Examination of the residual covariance matrix suggested that Time 2 and Time 3 covariances were more similar than either initial status to Time 2 or initial status to Time 3 covariances. This pattern indicates that individual's pseudoword performance after the onset of

instruction was more correlated than before instruction and after instruction performance.

Reading words from familiar story out of context. A model with time, treatment, and the Treatment × Time interaction as fixed effects, with a random effect for individual intercepts but not individual slopes, provided the best fit for the data, as indexed by smaller REML and larger AIC on the measure of reading words from familiar stories on word cards. A diagonal covariance structure that did not stipulate a pattern between covariances at each testing time fit the data best. The average intercept across children at initial status was significantly different from zero. Children averaged 7.86 ($SD = 1.60$) words read when the words from a familiar story were taken out of context. The random effect for intercept indicated that there were significant differences in children at initial status (Wald $Z = 4.86$, $p < .000$; see Table 3).

Reading words in context of familiar story. The fixed effects model, with time treated as a repeated measure, revealed that the intercept was significantly different from zero and that there was significant change in the number of words children read in context from Time 1 to Time 2 and from Time 2 to Time 3. However, in this case, the parameter estimates were negative. Children read 4.1 fewer words correctly at Time 2 and 1.56 fewer words correctly at Time 3. This probably was due to the very limited number of words and highly repetitive structure of the first story. The 26 test words for the first story were the entirety of words included in that story. There was no treatment or Treatment × Time interaction.

The model with a random effect that allowed intercepts and slopes to covary within individuals provided the best fit for the data, as indexed by smaller REML and larger AIC. Children who began reading more words in context showed higher growth rates than those who began reading fewer words (Wald $Z = 5.97$, $p < .000$). The heterogeneous autoregressive covariance (Wald $Z = 2.32$, $p < .02$) structure had an estimate of .58, indicating significant covariance of scores across time (see Table 3).

Reading unfamiliar words in context of unfamiliar story. The fixed effects estimates of the HLM analyses revealed a significant effect for the intercept and time but no effect for treatment or the Time × Treatment interaction. Children's average score at initial status was significantly greater than zero. From Time 1 to Time 2, children showed an average growth rate of 8.02 words, and from Time 2 to Time 3, children showed an average growth rate of 6.17 words. The model with a random effect that allowed intercepts and slopes to covary within individuals provided the best model fit and indicates that children began with significant differences in reading words in the context of an unfamiliar story, and these differences were related to rates of growth (see Table 3).

Additional word reading analyses. We wondered whether we may have missed treatment differences on word reading in text because these words had not been selected for having phonetic spellings (words in text and those same words on individual cards). Scores on the Time 3 reading words in context and out of context from the familiar story were reanalyzed with separate scores for the regular and irregular word spellings. There were more irregular than regular word spellings (19 words with irregular spellings and 7 words with regular spellings). Scores were transformed to percentages and then subjected to an arcsine transformation. These percentages were then compared between treatments with a multivariate analysis of variance (ANOVA). Although there was no significant mean differences between treatments on reading either regular or irregular words when they were embedded in a story or presented individually on word cards, there was a significant difference in variances between the two treatment groups on reading both phonetically regular and phonetically irregular words in context (Levene's test for equality of variances: regular words, $F[1, 54] = 6.28$, $p = .02$; irregular words, $F[1, 54] = 4.85$, $p = .03$). Children in the literature group had significantly greater variability in reading both types of words in the context of a familiar story than did children in the spelling group.

Across treatment groups, children also read a significantly larger percentage of the regular than irregular words both in context, $t(55) = 2.35$, $p = .02$, and out of context, $t(55) = 5.43$, $p < .001$. Thus, there was some indication that children in both groups with systematic phonics were developing enhanced word reading competencies for phonetically spelled words. This finding parallels the results for the spelling tasks in which overall children spelled phonetically regular words better than sight words (sight vs. real word spelling: $t[55] = 7.23$, $p < .001$; sight vs. pseudoword spelling: $t[55] = 6.06$, $p < .001$).

Paired samples t tests comparing February and May scores on reading words from familiar text stories in and out of context and comparing word reading in familiar and unfamiliar stories were performed. A significance level of .01 was adopted to adjust for the four comparisons that were made. These comparisons showed that across treatment groups and across two different stories (February and May), children read significantly more words when they were in the context of familiar stories than when the words were read in isolation (February: $t[55] = 11.83$, $p < .001$; May: $t[55] = 8.45$, $p < .001$) and that children read significantly more words in context from the familiar than the unfamiliar story in both February and May of first grade (February: $t[55] = 10.33$, $p < .001$; May: $t[55] = 8.32$, $p < .001$; see Table 2).

Writing

The HLM models with intercept, treatment, time of testing, and the Treatment × Time interaction specified as fixed effects, with different random effects (e.g., intercept, growth slope, intercepts plus slopes), did not converge. This means that these models did not adequately represent the structure of the data. Alternative models were then fitted to the data. Simplifying the model by eliminating the random effect and specifying a diagonal covariance structure, with time treated as a repeated measure, optimized the fit of the model, as indexed by smaller REML and larger AIC. This covariance structure indicates heterogeneous variances in writing fluency across time with no correlation between adjacent elements in the covariance matrix, indicating substantial individual fluctuation in writing fluency scores. The fixed effect for time was significantly different from zero, and the slopes for the two treatment treatments were significantly different on writing fluency. Average rates of growth in number of words written from Time 1 to Time 2 and from Time 2 to Time 3 were 28 and 24 words, respectively. This pattern shows a relatively stable rate of growth in writing fluency across first grade. The effect size in favor of the spelling group based on mean scores was .45 (see Table 3).

Correlations between both length of stories and words spelled correctly in written stories, and sight word spelling, phonetically regular real word spelling, and phonetically regular pseudoword spelling are shown in Table 4. These correlations show that sight

Table 4
End-of-First-Grade Correlations Among Writing Fluency, Spelling in Writing and Sight Word, Phonetic Real Word, and Phonetic Pseudoword Spelling

Measure	1	2	3	4	5
1. Writing fluency	—	.66***	.39**	.26	.16
2. Writing spelling		—	.76***	.44**	.32*
3. Sight word spelling			—	.45***	.42**
4. Phonetic real word spelling				—	.80***
5. Phonetic pseudoword spelling					—

* $p \leq .05$. ** $p \leq .01$. *** $p \leq .001$.

word spelling, words spelled correctly in written stories, and writing fluency are strongly and significantly correlated. Phonetic real word and phonetic pseudoword spelling are moderately and significantly correlated with correctly spelled words in written stories. This pattern suggests that treatment had its influence on writing fluency through the association between real and pseudoword spelling and the number of correctly spelled words in written stories.

Ability Effects on Reading, Spelling, and Writing Measures

Children were divided into three equal ability groups on the basis of their alphabet knowledge on entry to first grade. Three aspects of performance were evaluated by ability: mean performance on (a) reading, spelling, and writing; (b) oral reading of familiar and unfamiliar stories; and (c) word reading strategy during oral reading. A multivariate analysis, with treatment and ability as between-subjects factors, for the February and May reading, spelling, and writing measures was computed separately. The exact same pattern of results was found for both February and May. To simplify reporting, we computed the total score for Time 2 and Time 3 (scores sensitive to treatment effects) on the reading, spelling, and writing fluency measures, and we subjected them to a multivariate analysis, with treatment and ability as between-subjects factors. The multivariate effect for both treatment (Hotelling's $T^2 = 0.59$), approximate $F(1, 48) = 2.56$, $p = .02$, and ability (Hotelling's $T^2 = 2.62$), approximate $F(2, 48) = 5.64$, $p < .001$, was significant. There was no Ability × Treatment multivariate effect or any univariate Treatment × Ability effects (see Table 5).

There was a significant difference among all ability groups on the pseudoword reading, pseudoword spelling, real word spelling, familiar story reading, isolated reading of words from a familiar story, and word reading in an unfamiliar story measures (see Table 5). Children who entered first grade knowing the most letter sounds and letter names performed significantly better than both average- and low-ability children on letter names–sounds, and average children performed significantly better than low-ability children. On writing fluency and sight word spelling, high-ability children were significantly higher than both average- and low-ability children who did not differ, suggesting enhanced performance for the high-ability group on these two measures.

An analysis of the performance of the 18 children who were the lowest on letter name and letter sound knowledge at the beginning of first grade was conducted. The scores at the end of first grade for all 18 of these children on the spelling, reading, and writing fluency measures were arrayed and compared with the sample mean for each treatment group on these same measures. Across the 144 total scores for these 18 children on the three spelling, four reading, and one writing measure, only 32 scores (23%) met the sample average for the end of first grade. Of the 32 scores meeting the sample mean, 21 (66%) were achieved by children in the spelling-context group. There were 12 children (66%) who reached the sample mean on only one or two of the eight measures. Children who began first grade with low levels of letter–sound knowledge were highly likely to still be low at the end of first grade, regardless of which treatment they were in.

To examine children's in-context reading, we calculated the number and type of mispronunciations and the number of words on which no attempt was made during oral reading at the May testing. A word reading error was called a *phonetic mispronunciation* if the child's response indicated the use of phonic information (any phoneme in the word's spelling correctly pronounced) and was

Table 5
Oral Reading Means and Standard Deviations at the End of First Grade by Treatment and Ability

	Spelling								Literature							
	Low		Medium		High		Overall		Low		Medium		High		Overall	
Measure	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Phonetic miscues																
Familiar story	9.00	5.89	9.44	9.66	3.86	4.02	7.67	7.31	5.70	6.88	13.33	4.87	4.56	6.56	7.45	7.15
Unfamiliar story	5.75	4.62	3.22	3.70	1.00	1.15	3.42	3.91	4.40	4.54	3.33	2.83	0.75	1.36	2.68	3.37
Nonphonetic miscues																
Familiar story	2.13	2.36	1.88	2.26	0.57	0.79	1.58	2.01	3.10	4.20	3.56	3.24	1.08	3.45	2.45	3.70
Unfamiliar story	1.13	1.25	2.78	6.67	5.4	0.58	1.54	4.13	7.55	9.07	1.78	2.28	0.08	0.29	2.97	6.06
Strategy use																
Familiar story	3.77	1.09	4.44	1.81	5.85	0.38	4.60	1.55	2.50	1.43	4.67	1.11	5.92	0.29	4.45	1.77
Unfamiliar story	3.89	1.62	4.00	2.86	5.29	1.50	4.32	1.80	2.10	1.45	4.22	1.20	5.50	1.17	4.03	1.91
No attempt																
Familiar story	9.13	9.61	5.88	6.17	1.59	1.78	5.71	7.36	15.00	14.88	9.88	9.64	1.00	1.91	8.10	11.18
Unfamiliar story	8.75	9.27	3.56	4.36	0.57	0.79	4.28	6.55	11.50	10.17	3.56	3.21	0.58	1.44	4.97	7.55

counted as a nonphonetic error if no letter or combination of letters was given the correct pronunciation. It was hoped that these scores would elucidate children's decoding efforts while actually reading both a familiar and unfamiliar story. A multivariate analysis of these scores, with treatment and ability as between-subjects factors, yielded no treatment differences. A main effect for ability on the phonetic and nonphonetic miscues and number of words not attempted was found on both stories.

An estimate of the highest level of strategy use for word reading (range: 1–6) was made on the basis of an error analysis of each child's oral rendering of both the familiar and unfamiliar stories at Time 3. A score of 1 was given when a child told a story about the pictures or from recall in which there was little connection between the words in text and the words in the child's rendering. A score of 2 was given when the child showed an error pattern in which whole-word responses were made that were not based on the letters in words. Evidence of efforts to decode words on the basis of initial and or final phonemes was assigned a score of 3, whereas children who showed attempts to decode entire words were assigned a score of 4. Children who read words accurately with a fluency level of greater than 10 total hesitations were given a strategy score of 5, whereas those children who read words accurately and fluently with less than 10 total hesitations were given a strategy score of 6. Because this variable constituted an ordinal scale, we computed Mann–Whitney nonparametric tests for unrelated samples for each ability group. There was a significant effect for treatment in the low-ability group on strategy use scores on the unfamiliar (Mann Whitney $Z = 2.37$, $p = .02$) and familiar (Mann–Whitney $Z = 2.10$, $p = .04$) stories, but not in the middle-ability or high-ability groups (unfamiliar story: spelling, $M = 3.89$, $SD = 1.62$; literature, $M = 2.10$, $SD = 1.45$; familiar story: spelling, $M = 3.78$, $SD = 1.09$; literature, $M = 2.50$, $SD = 1.44$; see Table 4). Low-ability children in the spelling group relied more on decoding strategies during text reading than low-ability children in the literature group on both the unfamiliar and familiar story. The associated effect sizes were 1.16 and 1.00, respectively, indicating differences of at least one standard deviation.

First-Grade and Fifth-Grade Standardized Test Reading Comprehension

The effects of treatment on standardized reading comprehension were examined with two separate ANOVAs, with comprehension as the dependent variable and treatment as the independent variable. One analysis was computed for scores at the end of first grade on the MAT and another for scores at the end of fifth grade on the CAS. Both of these tests are nationally normed standardized tests (see Table 6). There was no significant difference between groups on comprehension scores at the end of first grade, but by the end of fifth grade, the spelling-context group scored significantly higher than the literature-context group on the standardized comprehension test, $F(1, 38) = 5.53$, $p = .02$. Raw scores were spelling context ($M = 18.00$, $SD = 3.52$) and literature context ($M = 15.39$, $SD = 3.69$). The national percentile scores were 70% and 50% for the spelling-context and literature-context groups, respectively. The effect size was 0.73, which is a moderately large effect. Fifth-grade comprehension scores were used because children moved from the elementary school to a Grade 6–8 middle school after fifth grade. There were 40 children remaining in the sample at the end of fifth grade, with 16 of the remaining children from the spelling-context group and 23 from the literature-context group. There were no significant group differences among the children remaining in the sample on any of the reading and spelling measures when they had begun first grade (see Table 2). To be cautious in attributing differences in fifth-grade comprehension to treatment that had occurred 4 years earlier, we calculated two covariate analyses. In one analysis, a composite score for individual differences in alphabet knowledge, spelling, and word reading at the beginning of first grade served as a covariate. In a second analysis, this same composite score plus end-of-first-grade comprehension scores were used as covariates. In these analyses, the effects of treatment on end-of-fifth-grade comprehension were amplified (see Table 6).

To determine which aspects of competence may have mediated the effect of treatment on fifth-grade comprehension scores, we performed a regression analysis. A composite of the pseudoword and real word spelling was computed to reduce the number of predictors in the model. End-of-first-grade phonetic word spelling, sight word spelling, pseudoword reading, and reading words from stories scores (3) were entered as independent variables. Spelling of phonetically regular words (pseudoword plus real word) and sight words at the end of first grade were significant predictors of fifth-grade comprehension. First-grade phonetic word spelling accounted for 8% of the variance in fifth-grade comprehension scores after removing variance associated with all other variables—including the intercorrelation among phonetic spelling and

Table 6
Fifth-Grade Comprehension Raw Scores and ANOVA Results With and Without Covariates

Analysis	Spelling treatment		Literature treatment					
	M	*SD*	*M*	*SD*	*F*	*dfs*	*p*	Partial η^2
Model 1								
Treatment	18.00	3.52	15.25	3.69	5.53	1, 38	.02	.13
Model 2								
Treatment					8.92	1, 37	.01	.19
Entry ability covariate					15.72	1, 37	.000	.20
Model 3								
Treatment					11.41	1, 36	.002	.24
Entry ability covariate					1.79	1, 36	.19	.05
End-of-first-grade comprehension covariate					3.92	1, 36	.06	.10

all other variables in the model. Sight word spelling also accounted for an independent 6% of the variance. This is a substantial influence when the conservative nature of the regression analyses is taken in to account.

Efforts to document that instruction in second grade to fifth grade did not advantage spelling-group children on comprehension were made. Accurate class lists were available for second grade and fifth grade. Two ANOVAs, with fifth-grade comprehension as the dependent variable and second-grade or fifth-grade class as independent variables, found no effect for class in either second or fifth grade, even though the chart presented in Table 7 shows that there was clear variability in reading instruction in both of these grades. In addition, 10 of the 12 second-grade to fifth-grade teachers were contacted and agreed to be retrospectively interviewed about their classroom literacy program and the procedures used for assigning children to classrooms at each grade level. Two teachers had left the district and could not be contacted. Teachers for all 10 of the classrooms independently reported the same schoolwide practice for assigning children to class. At the end of each year, current teachers created blue and pink cards for each of their current students. Overall ability and any significant behavioral issues were noted on these cards. Teachers at each grade level then created the class assignments for the next year teachers at a group meeting, with the goal of having the classes balanced for gender, ability, and significant behavior challenges. There was a built-in check for the equitableness of this procedure because teachers were making the class lists for their teacher colleagues. Teachers had used what was essentially a stratified random assignment approach for assigning children to classrooms each year. All but one of the interviewed teachers reported that this procedure resulted in evenly balanced classes. For the interviews, teachers were asked to describe their classroom practices, if any, for teaching phonics, reading comprehension, spelling, and writing. They were also asked to characterize their program as whole language, literature based, basal, or some combination of the three. Table 7 shows teacher responses by grade level.

Discussion

There are three major conclusions to be drawn from this study. First, instructional context influenced literacy outcomes, with the strongest treatment differences found on tasks most dependent on phonological competencies. Teaching phonics in the context of spelling was more effective than teaching phonics in the context of children's literature on measures of reading and spelling phonetically regular pseudowords and spelling phonetically regular real words. There was also evidence that these differences were maintained from the middle to the end of first grade and migrated to support the complex task of writing at the end of first grade and comprehension 4 years later. The effect sizes were moderate to large, ranging from 0.45 (writing) to over 1 (spelling and reading phonetically regular words). Finally, there were consistent differences in literacy performance across first grade among children with sample-based low, average, and high levels of letter–sound knowledge at the beginning of first grade. These differences were on the magnitude of 200%–300%. Low-alphabet-knowledge children who received phonics instruction embedded in literature fared less well than similar children in the spelling-context group. High-alphabet-knowledge children appeared to achieve relative areas of strength on measures not targeted in instruction including sight word spelling and writing fluency.

Instructional Context and Literacy Outcomes

Children in both treatments were equally successful in learning letter–sound correspondences. In February of first grade, the letter name mean for the literature group was 24.87 ($SD = 2.36$) and the mean for letter sounds was 20.32 ($SD = 3.53$). For the spelling group, the letter name and letter sound means were 25.19 ($SD = 2.00$) and 21.27 ($SD = 3.78$), respectively. In each group, children participated in 44 lessons on 23 letter–sound correspondences and encountered 8 exemplar words in each lesson for a total exposure to 368 words over the course of instruction. The high level of success in learning letter names (96%) and sounds (85%) for children in both groups with this moderate level of effort is noteworthy.

In contrast, teaching beginning phonics in the context of representing sounds in words with letters in partial spellings and attempting to spell simple words resulted in better performance for children in the spelling-context group on both reading and spelling words. Spelling-context children significantly outperformed the literature-context group in both February and May on measures of pseudoword reading and spelling. They also spelled real words significantly better in February. However, by the end of first grade (3 months after discontinuation of instruction), while spelling group scores were still higher, they were no longer significantly so as the rate of growth of spelling-context children decelerated after the discontinuation of instruction. Reading and spelling phonetically regular pseudowords is highly dependent on application of knowledge of letter–sound correspondences and segmenting and blending ability and is very unlikely to be accomplished by relying on memory for orthographic representations.

Performance on the reading and spelling of phonetically regular pseudowords and spelling phonetically regular real words was noticeably more modest compared with performance on the letter knowledge tasks as well (pseudoword reading performance in the spelling context was 65% and was 45% in the literature context; pseudoword spelling was 60% in the spelling context and 40% in the literature context; real word spelling was 65% in the spelling context and 50% in the literature context). Spelling instruction was also associated with greater use of decoding as a word reading strategy for children with low initial alphabet knowledge as they read both familiar and unfamiliar stories with uncontrolled vocabulary.

We suggest that enhanced utilization of phonological competencies derived from instruction underlies the superiority of the spelling treatment. Efforts to represent sounds in words decontextualized from text with partial- or whole-word spellings while learning letter–sound correspondences and blending apparently helped children apply their letter–sound knowledge to reading and spelling phonetically regular whole words. The spelling context produced better spelling and reading on those words drawing very systematically on encoding (spelling phonetically regular real and pseudowords) and decoding (reading pseudowords) of letter–sound correspondences. This initial advantage contributed to comprehension differences 4 years later.

What characteristics of the spelling-context treatment could have contributed to this enhanced application of letter–sound cor-

Table 7
Second-Grade to Fifth-Grade Teacher Interview Responses Describing Classroom Reading Comprehension, Phonics, Spelling, and Literature Instruction

Teacher	Type	Comprehension	Phonics	Spelling	Writing	Literature
			Second grade			
1	Whole language	Cloze technique, context cues, story maps	Decoding in text, minilessons (1 time per week), predictable texts, guided reading	Word dictation, sight word lists, weekly test	Daily writing, response journals, editing circles, sentence correction, writer's workshop	Class sets of books, reader's theater
2	Whole language	Poem of the week, pocket chart activities, context cues	Decoding in text, minilessons as needed, after school practice, leveled text	Word dictation, sight word lists, not formal lessons, weekly test	Daily writing, response journals, editing circles, writer's workshop	Class sets of books, listening centers
3	Literature-based basal, traditional	Factual questions on board	Decoding practice in workbook, ability groups with supplemental phonics for those who could not decode	Words from basal lists, sight words, story words, weekly test	Workbook pages asking for story reactions, story maps	Literature in basal, at end of year set of books used
			Third grade			
1	Whole language	Made stories from literature basal into plays, reader's theater, weekly poetry	None except choral reading	Homopones, words from poems, write words in sentence, weekly test	Daily writing, personal experiences, daily news, integrated with other subjects, not corrected	Classroom sets of books, lots of read-alouds
2	Literature-based basal	Activities from teacher's manual, workbook practice on short passages (answering questions)	Only what was in workbook and correction of individual children in individual reading	Word lists with basals; sight words, story words, maybe a spelling pattern or two	Journals for story responses and assigned writing topics, sentence practice	Literature in basal, library book reading
3	Literature-based basal, traditional	Factual questions from workbook, workbook pages on story analysis	After school help with the lowest readers, no program provided, helped with basic letter sounds	Basal word lists, practice and memory, weekly test	Story prompt written on board, story response probes in workbook	Classroom sets after basal completed
			Fourth grade			
1	Whole language	Integrated with social studies, socially important ideas emphasized, extended projects	None, except reading practice	Children made their child-selected lists of challenging important words, weekly test	Child-selected, long involved development of stories, personal journals, lots of writing	Personal books choices
2	Literature-based basal	Workbook activities, lots of interaction about story meaning	None	Basel spelling lists, weekly test	Workbook activities, special writing projects	Classroom sets of books with structured activities
3	—	—	—	—	—	—

Table 7 (*continued*)

Teacher	Type	Comprehension	Phonics	Spelling	Writing	Literature
			Fifth grade			
1	Literature-based basal	Book reports, activities from workbook, questioning, response to story	None	Word lists from basal word searches, crossword puzzles, weekly test	Journal writing, math writing, 4× weekly, kids ideas drove writing assignments, sentence correction	Class sets of book and "buddy" program with younger children for reading out loud
2	Literature-based basal	Prior knowledge, discuss the book with questions about the theme, read to students	None	Spelling lists not based on spelling patterns, but rather word themes and words from stories	Lots of journal writing, correction of sentence lessons, assigned topics, children selected topics	Read many sets of literature as a class, read books to children because of decoding problems
3	—	—	—	—	—	—

Note. Dashes indicate that data were not available.

respondence knowledge? An analysis of the cognitive operations involved in generating partial- and whole-word spellings leads to a number of possibilities. As spelling-group children assigned letters to either the initial or final position of words during letter–sound correspondence lessons, they had to segment individual sounds from word pronunciations and to concretely represent each sound with a selected letter. During whole-word lessons, children sequentially and systematically segmented sounds from spoken words, generated plausible letter associations for each of these sounds, and wrote these associations one at a time. Perhaps the actual act of writing letters may have enhanced children's ability to link orthography and phonemes, although children in each treatment learned letter–sound correspondences with equivalent levels of mastery. It is also possible that the need to write letters ensured children's attention during instruction.

Another possibility is that a focus on using phonics on individual, decontextualized words (rather than in text) may have improved phonetic spelling and word decoding by enhancing the salience of the letter–sound, segmenting, and blending instruction and activating task-relevant strategies. In the partial spelling exercises, attention was focused solely on representing initial and final letter sounds. The sequential sound-by-sound segmentation and sequential letter-by-letter conversion that occurred in the whole-word spelling lessons may have been more effective in engaging children in the thorough segmenting and blending necessary for accurate spelling and decoding than the equivalent amount of teacher-modeled segmenting and blending accompanied by children's oral responses and reading these practiced words in a literature excerpt that occurred in the literature-context treatment. The degree to which children's cognitive activity triggered by instruction makes the phonics competencies to be learned salient and/or ensures the deployment of task-relevant strategies (Siegler, 1996) may figure prominently in the acquisition of beginning reading phonics competence. Along this line of thinking, the degree of embeddedness makes a difference by influencing attention and strategy deployment. When phonics instruction is embedded in text, discourse meaning is most salient, and strategies best suited for comprehension are likely to be activated. In a disembedded approach, attention is more focused on the phonemic properties of individual words, and phonic strategies are more likely to be deployed. This idea has been referred to as *instructional press* by Roberts and Neal (2004). Even though we were not able to document the mechanisms for the spelling-context superiority, our task and instructional analyses implicate the plausible role of attention and/or strategy deployment.

The theoretically generated possibility that teaching phonics in the context of children's literature might lead to better reading in actual text or better writing on authentic writing tasks or better comprehension was not supported. The literature-context group, while equivalent to the spelling-context group on several measures, was not superior on any measure even when care had been taken to include measures that would be sensitive to potential benefits of each treatment.

It might be suggested that the measures of reading words in familiar and unfamiliar stories, in which there were no treatment effects, are more ecologically valid and thus more important. One likely reason there were no treatment differences on reading words from familiar stories both in and out of context is that before testing children had already read these stories 3–5 times. Consequently, children were very likely able to recognize words by either memory of the words in the story or the use of enriched context clues afforded by the repeated readings. The fact that children scored significantly better (1.5–2 times higher scores) on the words read in context than when they were presented in isolation is consistent with this interpretation. Another possibility is that reading the selected words whose phonological characteristics were not controlled may not have not been very sensitive to the influence of phonics instruction. The separate analysis of regular and irregular words on the familiar stories, which found no treatment effects, did not support this conjecture. However, there were only six words that were phonetically regular and all of these words had been read with guided reading a number of times. Children's word reading ability appeared very good when reading words in the context of familiar stories, an appearance not borne out by their context-free efforts on the same words.

The superiority of the spelling-context group was in comparison with an alternative, carefully matched systematic phonics program. Both the spelling-context group and the literature-context group experienced equivalent (a) amounts of instruction on the same sequence of letter–sound correspondences, (b) number of exemplars of each correspondence that was presented, (c) amount of blending and segmenting, (d) time allocated to the lessons, (e) requirement of an overt response to indicate whether target sounds occurred in the initial or final position of words, (f) segmentation of phonemes from whole words, and (g) cognitive activity with orthographic representations of sounds in words. Thus, time, instructional content and/or sequencing, overt responding, or exposure to orthographic representations of words cannot explain treatment effects. The two treatments differed primarily on whether the phonics skills were taught in the context of children's literature selections or in the context of efforts to spell individual words isolated from meaningful text. It is important to note that there was a rich variety of meaningful reading and writing activities in which children were encouraged to use their emerging letter–sound knowledge in the other components of the literacy program in place in both classrooms.

Children in the spelling context received very limited instruction in spelling principles beyond letter–sound correspondences but were instead asked to generate partial spellings to a significant degree. Consequently, there was less attention on complete and accurate orthographic representations than would be the case with conventional spelling instruction. Efforts to spell words conferred benefits on both reading and spelling words and suggests that encoding processes utilized during attempts to spell, in addition to any benefits derived from constructing memories for specific spellings and/or memory for spelling patterns, may be important in the early stages of reading.

These results also add to the existing evidence that the phonological processing involved in spelling and reading are interrelated as reported by other researchers (e.g., Ehri, 1989; Ehri & Robbins, 1992; Lundberg, Frost, & Petersen, 1988). Efforts to spell words, rather than instruction in spelling patterns or specific word spellings, generalized to pseudoword reading. Spelling and reading performance were also highly correlated (spelling context: $r = .71$; literature context: $r = .77$). Although there has been some debate regarding when spelling and reading become interrelated (Foorman, 1995), our data suggest that spelling and reading are linked as early as first grade if children are receiving phonics instruction.

Maintenance and Migration Effects

The overall pattern of significant differences between the spelling-context and literature-context groups was maintained between the end of treatment in February to the end of first grade, with the exception of correct spelling of phonetically regular real words. The Growth × Treatment interactions from Time 2 to Time 3 in the HLMs for the pseudoword reading and spelling tasks show that Time 2 to Time 3 growth—a time during which there was no further systematic phonics instruction—was significantly greater for spelling than literature-context children for both tasks even though the effect sizes were smaller after instruction was discontinued in February. The early benefits of the spelling context were sufficiently strong to maintain advantages for at least 3 months and were also a springboard for writing fluency. The divergence in growth trajectories for the two treatment groups on writing fluency, as measured by the number of words in stories seen at Time 2, had become statistically significant by Time 3. This finding is compatible with the interpretation that explicit instruction of the alphabetic principle in the context of spelling served as a scaffold for meaningful expressive writing. In other words, promoting facility with the word-level skill of spelling supported the more complex process of expressive writing. This result replicates those of Berninger et al. (1998) and of Graham, Harris, and Chorzempa (2002) and extends them by showing that the link between spelling and writing fluency is evident by the end of first grade.

Correlations among writing fluency, correct spelling in writing, and performance on spelling tests showed that indeed phonetic real word and phonetic pseudoword spelling were moderately but significantly correlated with correctly spelled words written in stories, which was then significantly associated with writing fluency ($r = .66$) and phonetic word spelling was influenced by treatment. However, this explanation is not entirely satisfying because children in the spelling group did not accurately spell more words in their writing when we counted either completely accurate word spellings or developmental spellings in which credit was given for phonetically plausible partial spellings. The strong correlation ($r = .76$) between sight word spelling (on which there was no treatment effect) and correctly spelled words in stories suggests that many of the words children used in stories were phonetically irregular words.

The initial benefits in spelling and reading words associated with phonics instruction delivered in the context of generating spellings migrated to comprehension 4 years later. The advantage was moderately large, as indicated by an effect size of 0.64. This pattern is consistent with the interpretation that children in the spelling group had an early advantage in applying letter–sound correspondence knowledge to word reading and spelling that with time migrated to facilitating comprehension processes. Shankweiler et al. (1999) showed word-level effects on comprehension up through second grade. This study extends those results by showing the influence of word reading skill on comprehension at the end of fifth grade. The fifth-grade comprehension result is not an artifact of selective attrition in the two treatments because performance of children remaining at the end of fifth grade (a) was equivalent at the beginning of first grade and (b) showed the same pattern of treatment effects at the end of first grade, as reported for the initial full sample. Although efforts were taken to document that instruction in the intervening years was unlikely to have advantaged spelling-group children, the indirect nature of this evidence for third and fourth grades indicates that these results should be interpreted cautiously.

It was not only phonological abilities that contributed to end-of-fifth-grade comprehension. The regression analysis showed that sight word spelling and spelling phonetically regular words (pseudo and real) were the two independent predictors of fifth-grade comprehension. Morris, Bloodgood, and Perney (2003) reported a similar finding. They found that spelling with beginning and ending consonants was a better predictor of reading 2.5 years later than an oral segmenting task. In the present study, sight word reading was not significantly influenced by treatment. Both the ability to segment and orthographically represent speech (phonetic spelling) and the ability to remember orthographic representations for words (sight words) were predictive of later comprehension.

The importance of sight word spelling was also found on the writing measure.

There was a clear floor effect on the first-grade MAT, providing at least a partial accounting for failing to find any comprehension differences between treatments at the end of first grade. Although the data do not identify the reasons for this poor performance on the MAT, the great difficulty children had with understanding how to take the test and their inability to attend to and attempt the more difficult items surely contributed. Twenty-four children did not complete the comprehension section of the test. In accounting for delayed effects on comprehension of their spelling intervention, Uhry and Shepherd (1993) suggested that oral comprehension is influenced before silent comprehension.

Ability Effects

A number of studies consistently reveal that children, particularly those most at risk for reading failure, benefit from greater amounts of explicit instructional support for learning phonemic awareness and the alphabetic principle (e.g., Becker & Gersten, 1982; Blachman, Tangel, Ball, Black, & McGraw, 1999; Bowey, 1995; Foorman et al., 1998; Torgeson et al., 1999). Others have suggested that a meaning emphasis and use of children's previous knowledge and experience, which is characteristic of many holistic emphasis programs, should be more beneficial for lower socioeconomic status children. We found that more decontextualized phonics instruction was more efficacious than matched phonics instruction that was embedded in children's literature for children at risk for reading difficulties on the basis of both familial socioeconomic status (participating children were predominantly from families with below middle-class socioeconomic levels) and ability groupings empirically derived from beginning-of-first-grade letter name and letter–sound knowledge.

Our findings also provide some limited evidence for the enhanced effectiveness of the spelling-context group for children with initial low alphabet knowledge. These children had less developed decoding strategies for reading words in stories compared with children with middle or high initial alphabet knowledge. Only 32 (23%) of the 144 total end-of-first-grade reading and spelling scores for children with low knowledge at the beginning of first grade reached the sample mean. Sixty-six percent of these 32 scores were achieved by the low-ability children in the spelling-context group.

Regardless of which treatment they were in, children with low initial alphabet knowledge were behind high alphabet knowledge children on every measure and behind average alphabet knowledge children on almost every measure at the three testing points. Most of these children remained markedly low across the first-grade year. The HLM analyses amplify these ability results. On all three of the reading and spelling tasks that relied the most on phonological processing, children who started higher tended to have stronger growth rates. Low-alphabet-knowledge children even performed significantly lower than average- and high-alphabet-knowledge children on reading words in context from a story that had been substantially scaffolded through repeated and choral readings with the teacher and peer partners. This rather sobering profile for children who entered first grade with low alphabet knowledge points out that the letter–sound and blending instruction included in this program were not sufficient for children most at risk for reading difficulty on the basis of low alphabet knowledge at entry to first grade. Instruction across the entire first-grade year with more work on spelling patterns, review and practice of letter–sound correspondences, more sophisticated blending, and efforts to target for more intensive intervention those children showing the least growth was probably needed (O'Connor, 2000).

Implications for Instruction

First, instruction utilizing spelling as a context for teaching grapheme–phoneme correspondence, blending, and segmenting had a significantly greater positive influence on both reading and spelling than teaching phonics in the context of literature and should therefore be given strong consideration as an instructional context of choice for systematic phonics instruction. Although this study documents that phonics can be taught in the context of literature selections and that this approach was as effective for several measures of reading performance as teaching phonics through spelling, it was difficult to find stories with sufficient numbers of words for several of the letter–sound correspondences. Teachers must be very knowledgeable and prepared to invest the necessary time and effort to find literature that lends itself to systematic instruction. A recent report by Cunningham, Perry, Stanovich, and Stanovich (2004) indicates that it may be challenging for teachers to meet this expectation. They reported that teacher's knowledge of both literature and phonics was not comprehensive and that teachers overestimated their knowledge in both areas. In the present study, the effectiveness of the literature context occurred when there was a very systematic and explicit phonics component connected to the literature selection on every day of instruction. If teachers elect to systematically teach phonics through the use of children's literature, it is important to supplement this approach with instruction in using phonics knowledge to generate spellings for isolated words.

An instructional context in which phonics instruction was disembedded from meaningful text promoted reading and spelling of phonetically regular words. It also helped the more complex processes of writing and comprehension. There was no evidence that children's beginning word reading and spelling—in either decontextualized or connected text—was negatively impacted by disembedded phonics instruction when the complete classroom literacy program included extensive opportunity to engage with meaningful and authentic text and writing activities. Efforts to increase children's sight word spelling appears important given its relationship with writing fluency and comprehension. The differential ease of learning letter–sound correspondences compared with word decoding and word spelling suggests that the ratio of instruction between these two elements may need to reflect this disparity. The added value of instruction that explicitly helped children generate partial- and whole-word spellings using their emerging phonics knowledge is important. Children in both groups engaged in daily expressive writing in which invented spelling was encouraged during the nonphonics part of classroom literacy activity. Those children who received instruction and structured practice on how to do invented spelling acquired and maintained word reading, word spelling, and story writing advantages.

Teachers who work with economically disadvantaged children and teach in high student:teacher ratio classrooms effectively taught systematic phonics in the context of spelling, and this

instruction fostered children's use of their phonic knowledge in spelling, reading, and writing. These benefits were maintained and benefited later reading comprehension. This study has shown that small variations in instructional context for teaching phonics to socioeconomically at-risk children matter. More research exploring how other specific, theoretically generated variations in instructional context influence the acquisition and application of phonics knowledge is needed.

References

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

Adams, M. J. (1998). The three-cueing system. In F. Lehr & J. Osborn (Eds.), *Literacy for all: Issues in teaching and learning.* New York: Guilford Press.

Ball, E., & Blachman, B. (1988). Phonological segmentation training: Effects on reading readiness. *Annals of Dyslexia, 38,* 208–225.

Becker, W., & Gersten, R. (1982). A follow-up of Follow Through: The later effects of the Direct Instruction Method on children in fifth and sixth grades. *American Educational Research Journal, 19,* 75–92.

Bergeron, B. (1990). What does the term whole language mean? Constructing a definition from the literature. *Journal of Reading Behavior, 22,* 301–323.

Berninger, V. W., Vaughan, K., Abbott, R. D., Begay, K., Coleman, K. B., Curtin, G., et al. (1998). Teaching spelling and composition alone and together: Implications for the simple view of writing. *Journal of Educational Psychology, 94,* 291–304.

Blachman, B., Tangel, D., Ball, E., Black, R., & McGraw, D. (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,* 273–293.

Bowey, J. A. (1995). Socioeconomic status differences is preschool sensitivity and first-grade reading. *Journal of Educational Psychology, 87,* 476–487.

Bradley, L., & Bryant, P. E. (1983, February 3). Categorizing sounds and learning to read—A causal connection. *Nature, 301,* 419–421.

Byrne, B., & Fielding-Barnsley, R. (1989). Phonemic awareness and letter knowledge in the child's acquisition of the alphabet principle. *Journal of Educational Psychology, 81,* 313–321.

California State Department of Education. (1996). *Recommended literature: Kindergarten through grade twelve*. Sacramento, CA: Author.

Center, Y., Freeman, L., & Robertson, G. (2001). The relative effect of a code-oriented and meaning-oriented early literacy program on regular and low progress Australian students in Year 1 classrooms which implement reading recovery. *International Journal of Disability, Development and Education, 48,* 207–232.

Chall, J. S. (1983). *Learning to read: The great debate.* New York: McGraw-Hill.

Chomsky, C. (1979). Approaching reading through invented spelling. In L. B. Resnick & P. A. Weaver (Eds.), *Theory and practice of early reading* (Vol. 2, pp. 43–64). Hillsdale, NJ: Erlbaum.

Christensen, C. A., & Bowey, J. A. (2005). The efficacy of orthographic rime, grapheme–phoneme correspondence, and implicit phonics approaches to teaching decoding skills. *Scientific Studies of Reading, 9,* 327–349.

Clarke, L. K. (1988). Invented spelling versus traditional spelling in first graders' writings: Effects on learning to read and spell. *Research in the Teaching of English, 22,* 281–309.

Cunningham, A. E., Perry, K. E., Stanovich, K. E., & Stanovich, P. (2004). Disciplinary knowledge of K-3 teachers and their knowledge calibration in the domain of early literacy. *Annals of Dyslexia, 54,* 139–165.

Ehri, L. C. (1997). Learning to read and learning to spell are one and the same, almost. In C. A. Perfetti, L. Rieben, & M. Fayol (Eds.), *Learning to spell: Research, theory and practice across languages* (pp. 237–269). Mahwah, NJ: Erlbaum.

Ehri, L. (1989). Movement into word reading and spelling: How spelling contributes to reading. In J. M. Mason (Ed.), *Reading and writing connections* (pp. 45–81). Boston: Allyn & Bacon.

Ehri, L. C. (1992). Reconceptualizing the development of sight word reading and its relationship to recoding. In P. B. Gough, L. C. Ehri, & R. Treiman (Eds.), *Reading acquisition* (pp. 107–144). Hillsdale, NJ: Erlbaum.

Ehri, L. C., & Robbins, C. (1992). Beginners need some decoding skills to read words by analogy. *Reading Research Quarterly, 27,* 12–26.

Ehri, L. C., & Wilce, L. (1980). Do beginners learn to read function words better in sentences or lists? *Reading Research Quarterly, 15,* 451–476.

Ehri, L. C., & Wilce, L. (1987a). Cipher versus cue reading: An experiment in decoding acquisition. *Journal of Educational Psychology, 79,* 3–13.

Ehri, L. C., & Wilce, L. (1987b). Does learning to spell help beginners learn to read words? *Reading Research Quarterly, 22,* 47–65.

Elley, W. B. (1989). Vocabulary acquisition from listening to stories. *Reading Research Quarterly, 24,* 174–187.

Foorman, B. R. (1995). Research on the "great debate": Code-oriented versus whole language approaches to reading instruction. *School Psychology Review, 24,* 376–392.

Foorman, B. R., & Francis, D. J. (1994). The relevance of a connectionist model of reading for "the great debate." *Educational Psychology Review, 6,* 25–47.

Foorman, B. R., Francis, D. J., Fletcher, J. M., Schatshneider, 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.

Foorman, B. R., Francis, D. J., Novy, D. M., & Liberman, D. (1991). How letter–sound instruction mediates progress in first-grade reading and spelling. *Journal of Educational Psychology, 83,* 456–469.

Goodman, K. (1986). *What's whole in whole language?* Ontario, CA: Scholastic.

Gough, P. B., Juel, C., & Griffith, P. (1992). Reading, spelling and the orthographic cipher. In P. B. Gough, L. C. Ehri, & R. Treiman (Eds.), *Reading acquisition* (pp. 36–38). Hillsdale, NJ: Erlbaum.

Graham, S. (2000). Should the natural learning approach replace spelling instruction? *Journal of Educational Psychology, 92,* 235–247.

Graham, S., Harris, K. R., & Chorzempa, B. F. (2002). Contribution of spelling instruction to the spelling, writing, and reading of poor spellers. *Journal of Educational Psychology, 94,* 669–686.

Gunderson, L., & Shapiro, J. (1988, January). Whole language instruction: Writing in 1st grade. *The Reading Teacher,* 430–436.

Guthrie, J. T., Wigfield, A., & VonSecker, C. (2000). Effects of integrated instruction on motivation and strategy use in reading. *Journal of Educational Psychology, 92,* 331–341.

Harris, A. J., & Sipay, E. R. (1990). *How to increase reading ability*. New York: Longman.

Holdaway, D. (1982). Shared book experience: Teaching reading using favorite books. *Theory into Practice, 21,* 293–300.

International Reading Association and National Association for the Education of Young Children. (1998, July). Learning to read and write: Developmentally appropriate practices for young children. *Young Children,* 30–46.

Lundberg, I., Frost, J., & Petersen, O. (1988). Effects of an extensive program for stimulating phonological awareness in preschool children. *Reading Research Quarterly, 23,* 263–284.

McCracken, M. J., & McCracken, R. A. (1982). *Spelling through phonics*. Manitoba, Canada: Peguis.

McGuiness, D., McGuiness, C., & Donohue, J. (1995). Phonological training and the alphabet principle: Evidence for reciprocal causality. *Reading Research Quarterly, 30,* 830–852.

Morris, D., Bloodgood, J., & Perney, J. (2003). Kindergarten predictors of

first grade reading achievement. *Elementary School Journal, 104,* 93–111.

Morrow, L. M. (1992). The impact of a literature-based program on literacy achievement, use of literature and attitudes of children from minority backgrounds. *Reading Research Quarterly, 27,* 251–275.

Neumann, S. B. (1999). Books can make a difference: A study of access to literacy. *Reading Research Quarterly, 34,* 286–311.

Newman, J. M., & Church, S. M. (1990). Myths of whole language. *The Reading Teacher, 44,* 20–26.

NICHHD. (2000). *Report of the National Reading Panel: Teaching children to read. An evidence-based assessment of the scientific research on reading and its implications for reading instruction* (NIH Publication No. 00–4769). Washington, DC: U.S. Government Printing Office.

O'Connor, R. (2000). Increasing the intensity of intervention in kindergarten and first grade. *Learning Disabilities Research and Practice, 15,* 43–54.

O'Connor, R., & Jenkins, J. (1995). Improving the generalization of sound/symbol knowledge: Teaching spelling to kindergarten children. *Journal of Special Education, 29,* 255–275.

Pemberton, E. F., & Watkins, R. V. (1987). Language facilitation through stories: Recasting and modeling. *First Language, 7,* 79–89.

Perfetti, C., & Hogaboam, T. (1975). Relationship between single word decoding and reading comprehension skill. *Journal of Educational Psychology, 67,* 461–469.

Richgels, D. (1995). Invented spelling ability and printed word learning in kindergarten. *Reading Research Quarterly, 30,* 96–109.

Robbins, C., & Ehri, L. C. (1994). Reading stories to kindergartners helps them learn new vocabulary words. *Journal of Educational Psychology, 86,* 54–64.

Roberts, T., & Neal, H. (2004). Relationships among preschool English language learner's oral proficiency in English, instructional experience and literacy development. *Contemporary Educational Psychology, 29,* 283–311.

Routman, R. (1994). *Invitations: Changing as teachers and learners K-12.* Portsmouth, NH: Heinemann.

Senechal, M., & Cornell, E. H. (1993). Vocabulary acquisition through shared reading experiences. *Reading Research Quarterly, 28,* 123–138.

Senechal, M., Thomas, E. H., & Monker, J. A. (1995). Individual differences in 4-year-old children's acquisition of vocabulary during storybook reading. *Journal of Educational Psychology, 87,* 218–229.

Shankweiler, D., Lundquist, E., Katz, L., Stuebing, K. K., Fletcher, J. M., Brady, S., et al. (1999). Comprehension and decoding: Patterns of association in children with reading difficulties. *Scientific Studies of Reading, 3,* 69–94.

Siegler, R. S. (1996). *Emerging minds: The process of change in children's thinking.* New York: Oxford University Press.

Silva, C., & Martins, M. A. (2003). Relations between children's invented spelling and the development of phonological awareness. *Educational Psychology, 23,* 3–16.

Stahl, S., Duffy-Hester, A. M., & Stahl, K. A. (1998). Everything you wanted to know about phonics (but were afraid to ask). *Reading Research Quarterly, 35,* 338–355.

Stahl, S., & Murray, B. (1998). Issues involved in defining phonological awareness and its relation to early reading. In J. L. Metsala & L. C. Ehri (Eds.), *Word recognition in beginning literacy* (pp. 65–85). London: Erlbaum.

Strickland, D. (1998). *Teaching phonics today: A primer for educators.* Newark, DE: International Reading Association.

Tangel, D. M., & Blachman, B. A. (1995). Effect of phoneme awareness instruction on the invented spelling of first grade children: A one year follow-up. *Journal of Reading Behavior, 27,* 153–185.

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

Treiman, R. (1998). Why spelling? The benefits of incorporating spelling into beginning reading Instruction. In J. L. Metsala & L. C. Ehri (Eds.), *Word recognition in beginning literacy* (pp. 289–313). London: Erlbaum.

Tunmer, W. E., Herriman, M. L., & Nesdale, A. R. (1988). Metalinguistic abilities and beginning reading. *Reading Research Quarterly, 23,* 134–158.

Tunmer, W. E., & Nesdale, A. R. (1985). Phonemic segmentation skill and beginning reading. *Journal of Educational Psychology, 77,* 417–427.

Uhry, J. K., & Shepherd, M. J. (1993). Segmentation/spelling instruction as part of a first grade reading program: Effects on several measures of reading. *Reading Research Quarterly, 28,* 219–233.

Weaver, C. (1994). *Understanding whole language: From principle to practice* (2nd ed.). Portsmouth, NH: Heinemann.

Yopp, H. K. (1988). The validity and reliability of phonemic awareness tests. *Reading Research Quarterly, 23,* 153–185.

(*Appendixes follow*)

Appendix A

Words for the Three Spelling Lists

	Real words			Pseudowords		
Sight words	List 1	List 2	List 3	List 1	List 2	List 3
Was	Lab	Lit	Jab	Bal	Til	Baj
Say	Tim	Tom	Dim	Tim	Mot	Mid
Want	Slob	Slot	Slim	Slad	Slif	Slag
Does	Fob	Cob	Dot	Bof	Boc	Tod
Come	Bit	Sit	Fit	Tib	Tis	Tif
Said	Tan	Tag	Nag	Nat	Gat	Gan
What	Scat	Skim	Skit	Scif	Skap	Skom
Have	Mop	Sop	Cop	Pom	Pos	Poc
Went	Cad	Sad	Fad	Dac	Das	Daf
Play	Spin	Spat	Spot	Spod	Spof	Spic
Here	Cod	Pod	Sod	Doc	Dop	Dos
They	Din	Fin	Bin	Nid	Nif	Nib
You	Sam	Jam	Man	Mas	Maj	Nam
Them	Jog	Bog	Fog	Goj	Gob	Gof
Saw	Crib	Crop	Crab	Crot	Crim	Crod
Could	Fib	Bid	Sin	Bif	Dib	Nis
Again	Sob	Pad	Mob	Bos	Dap	Bom
Because	Fan	Fat	Sat	Naf	Taf	Tas
Found	Pig	Fig	Dig	Gip	Gif	Gid
When	Prom	Prim	Prod	Prit	Prot	Prit

Appendix B

Word Lists for Familiar and Unfamiliar Text Stories

Familiar story (in context and on word cards)			Unfamiliar story (in context only)		
October	February	May	October	February	May
His	Teach	Need	The	Come	Creaked
Pink	Store	Store	Boat	In	Blew
Who	Me	Will	Go	Like	Said
Bandana	Every	First	Will	Bear	Long
Brown	Own	Wanted	You	Sad	Houses
Jeans	Another	Bed	Help	Duck	Name
Blue	Dinner	Dogs	Hen	Robot	Kettle
Sweater	This	Look	Yes	Time	Hiss
Violet	Clean	Sat	Frog	Lunch	There
Hat	Pet	Gray	Cat	For	Roof
Purple	Little	Sign	Bear	Bring	Fell
Ribbons	Throw	Must	Won't	What	Old
Shirt	Tail	Hard	It	Asked	Tea
Red	Tricks	Soon	Told	Anytime	Wind
Her	Scratch	Buy	Get	Just	Noisy
Were	One	Made	So	This	Trees
Night	Puppy	Sell	To	Fun	Bed
Long	Black	Better	I	He	Man
Day	Soon	Went	No	Going	Swish
Sneakers	Will	Doghouse	Do	Nice	Leaves
Yellow	Live	Someone	I'll	Work	Outside
Green	Kid	Worked		Make	Time
Orange	Pulls	Paintbrush		Nice	Lived
All	Comes	Sale		Us	Whistled
Dress	Kisses	One		Going	Though
Pants	Right	Come		Get	Floor

Appendix C

List of Literature Selections Used in the Phonics in Context Treatment

Title	Author
1. *Mike Mulligan and His Steam Shovel*	Virginia Lee Burton
2. *Stone Soup*	Marcia Brown
3. *One Fine Day*	Nonny Hogrogian
4. *Blueberries for Sal*	Robert McCloskey
5. *Dandelion*	Don Freeman
6. *Ten, Nine, Eight*	Molly G. Bang
7. *The Little Engine that Could*	Watty Piper
8. *The Very Busy Spider*	Eric Carle
9. *Corduroy*	Don Freeman
10. *Little Red Hen*	Paul Galdone
11. *The Tenth Good Thing about Barney*	Judith Viorst
12. *Mr. Rabbit and the Lovely Present*	Charlotte Zolotow
13. *Freight Train*	Donald Crews
14. *The Tale of Peter Rabbit*	Beatrix Potter
15. *The Cat in the Hat*	Dr. Seuss
16. *The Happy Lion*	Louise Fatro
17. *Ask Mr. Bear*	Marjorie Flack
18. *Momo's Kitten*	Mitsu & Taro Yashima
19. *Strega Nona*	Tomie dePaola
20. *Caps for Sale*	Esphyr Slobodkina
21. *Bread and Jam for Francis*	Russell Hoban
22. *Make Way for Ducklings*	Robert McCloskey
23. *Harry and the Terrible Whatzit*	Dick Gackenbach
24. *Harry the Dirty Dog*	Gene Zion
25. *Petunia*	Roger Duvoisin
26. *Frog and Toad are Friends*	Arnold Lobel
27. *Alexander and the Wind-Up Mouse*	Leo Lionni
28. *Bringing the Rain to Kapiti Plain*	Verna Aardema
29. *William's Doll*	Charlotte Zolotow
30. *Father Fox's Pennyrhymes* ("Miss Quiss")	Clyde Watson
31. *Once a Mouse*	Marcia Brown
32. *Big Bad Bruce*	Bill Peet
33. *Umbrella*	Taro Yashima
34. *The Shoemaker and the Elves*	Jakob & Wilhelm Grimm

Received May 20, 2005
Revision received June 13, 2006
Accepted June 20, 2006 ■