A Comparison of First Graders' Reading Acquisition with Little Books and Literature Anthologies

Shailaja Menon

University of Colorado, Boulder

Elfrieda H. Hiebert

University of Michigan

Direct correspondence regarding this manuscript to:

Elfrieda H. Hiebert
hiebert@umich.edu

or:
106 Phelan Court
Santa Cruz, CA 95060

Running Head: Reading Acquisition with Little Books and Anthologies

Abstract

This study examined the effectiveness of a little book curriculum in facilitating the independent reading skills of first-grade readers. The curriculum was based on a theoretical model that identified two critical dimensions of text-based support for beginning readers: linguistic content and cognitive load.

The 15-week little book intervention was conducted in four first-grade classrooms of an inner-city school that was part of a large-city school district. Two of the classes were assigned to the intervention group and the other two classes were assigned to the comparison group. Children in the intervention group read from little books leveled according to features of linguistic content and cognitive load. Children in the comparison group read from basal literature texts. Word lists and graded passages from the Qualitative Reading Inventory (QRI) served as the pre- and post-test measures. ANCOVA and chi-square analyses revealed that children in the intervention classrooms performed at significantly higher levels on the post-tests than their counterparts in the comparison classrooms. These results applied equally to the word-lists and the passage reading tasks. The intervention was effective with children at all reading levels--high, average and low.

A Comparison of First Graders' Reading Acquisition with Little Books or Literature Anthologies

The study is based on the assumption that texts help define the parameters of beginning readers' success in reading. Texts are by no means the only important element in the reading acquisition process. As Barr (1972) and Juel and Roper/Schneider (1985) have reported, instructional method and text combine to shape children's word recognition strategies. The less consistent the features of text, the more dependent beginning readers are on instructional method and, in turn, their teacher's efficacy in employing the instructional method. At a time when the teaching force is in transition, texts are a critical component of beginning reading instruction, especially for children in economically challenged schools.

Over the past two decades, beginning reading texts have undergone substantial changes. The majority of beginning reading texts in American schools emphasized high-frequency words through the mid-1980s (Anderson, Hiebert, Scott & Wilkinson, 1985). By the mid-1990s, the beginning reading texts of major programs consisted of predictable and literary texts (Hoffman, McCarthey, Abbott, Christian, Corman, Curry, Dressman, Elliot, Maherne, & Stahle, 1994). By 2002, the content of beginning textbooks had changed again. Responding to the mandates for the Texas's adoption of texts in 2000 and those of California in 2002, all mainstream basal programs now offer decodable texts in their beginning reading components (Hoffman, Sailors, & Patterson, 2002).

While each change in beginning textbooks is accompanied by claims of research-based evidence, policy-makers have been a primary source for these recent changes. In making the rapid changes mandated by policy-makers, publishers have been left to make choices about features of texts that can influence beginning reading acquisition but have not been addressed by

policy-makers. The swift changes in textbook features have not been based on comprehensive models of the critical text features that influence reading acquisition.

This study fills this gap by testing a model of text features on first graders' acquisition of independent word-solving skills, which strongly predict independent reading skills (Juel, Griffith &Gough, 1986; Shanahan, 1984). The model that underlies this study, the Text Elements by Task (TExT) model, identifies two critical dimensions of text-based support for independent reading: cognitive load and linguistic knowledge. In this study, the explanatory efficacy of this model was empirically examined through the progress toward independent reading skills made by two groups of first graders. One group received the literature anthologies of basal reading programs that continue to be the mainstay of American reading instruction (Baumann, Hoffman, Moon, & Duffy-Hester, 1998). The other group read from short texts called little books that were ordered according to the TExT model and matched to children's reading levels.

The view that texts scaffold the beginning reading process is by no means controversial. Views about the particular features of texts that scaffold the beginning reading process are more controversial. To ground the TExT model in past and current perspectives of the scaffolding provided by text, the primary perspectives are reviewed. This overview is followed by a description of the theoretical and empirical basis for the TExT model and, subsequently, the questions addressed in this study.

Views on Texts as Scaffolds

Three prominent views on the features of text that scaffold the task for beginning readers can be identified within scholarship and practice: (a) pacing and repetition of high-frequency words, (b) predictable text structures and engaging content, and (c) decodable elements.

Pacing and repetition of words. Beginning with the 1930 edition of the Scott Foresman readers (Elson & Gray, 1930), William S. Gray applied two components of Thorndike's work to the beginning reading texts that came to be known as the "Dick and Jane" texts, referring to two of the texts' characters. The model's first component was its concentration on high-frequency words, following Thorndike's (1921) analyses of the frequency of word appearance in English texts. The model's second component drew on Thorndike's (1903) four laws of learning—effect, exercise, readiness, and identical elements. These elements of strict control of new words applied to the first four textbooks of the first-grade program: the three pre-primers and the primer. By the time that students got to the last first-grade book—the first-grade reader—Dick and Jane and their friends and family had been replaced with a variety of characters and the pacing and repetition of new words had been replaced with readability formulas. In the readability formulas that were used in most textbook programs (Spache, 1974; Chall & Dale, 1995), the vocabulary was measured against particular groups of common vocabulary. However, readability formulas did not attend to the formulaic repetition of words as Gray and his colleagues did in writing first-grade texts (Elston & Gray, 1930; Gray, Monroe, Artley, Arbuthnot, & Gray, 1956).

Sentence and text structures. In addition to assessing the match of the words in a text with a prescribed vocabulary, readability formulas assess sentence length. Sentences were kept short in the Dick and Jane texts but the designers of the early reading texts did not overtly address sentence length itself. However, at later levels, complex sentences were often broken into separate simple sentences to make texts comply with the readability requirements of particular grade levels. In the late 1970s and early 1980s, the effects of such manipulations in texts on student comprehension became the focus of numerous studies (e.g., Brennan, Bridge &

Winograd, 1986). The youngest students in these studies were second graders (Brennan et al., 1986), a level controlled by readability formulas, not by Gray's formula for beginning texts. The findings from these studies, however, were applied to controlled text of any sort. In Becoming a Nation of Readers (Anderson et. al, 1985), the national report that summarized the findings on text manipulations for educators and policy-makers, examples from primers were used to illustrate controlled texts.

The policy-makers in America's two largest states that also conduct state-wide textbook adoptions agreed with Becoming a Nation of Readers' call for an end to controlled texts. In California's textbook mandate of 1987 (California English/Language Arts Committee, 1987), the mandates for authentic texts or literature extended from first grade through high school. The 1990 guidelines of Texas (Texas Education Agency, 1990) called for similar texts. Recognizing that beginning readers required some form of scaffolding, publishers offered predictable texts as the form of authentic texts in their beginning reading components. In predictable texts, a phrase, sentence, or even group of sentences is repeated to form the text structure, as in This is the house that Jack built. The underlying philosophy is that the cadence and the repetition of the predictable syntactic and textual patterns permit beginning readers to read along with a capable reader from their initiation into reading and, after rereading the text by themselves, to develop sight vocabularies (Goodman, 1968).

Research on whether predictable sentence and text structures scaffold beginning reading recognition was limited in scope when predictable texts became commonplace in mainstream textbook programs. While two studies by Bridge and colleagues (Bridge & Burton, 1982; Bridge, Winograd, & Haley, 1983) were cited as evidence that sight words were learned through predictable texts, other studies with more extended samples and rigorous techniques indicated

that beginning and challenged readers relied on aural memory when reading predictable texts (Leu, DeGroff, & Simons, 1986). More recently, Johnston (2000) has shown that even the most proficient first graders learn only a fraction of the words in predictable texts. Over-reliance by beginning readers on predictable syntactic and text structures has the same effect as over-reliance on illustrations in that children fail to develop strong grapho-phonic strategies (Samuels, 1970).

Instructed decodable elements. While the model of pacing and repetition of a particular group of words was primarily applied with high-frequency words, the same principles of pacing and repetition were also applied to phonetically regular words. For example in the Scribner Reading Program (Cassidy, Roettger, & Wixson, 1987), the phonetically regular word "run" is introduced in one text and used seven times. In the next passage, it appears five times. These texts never attained the prominence, however, of those that used the high-frequency words, even after Chall's (1967/1983) critique of the mainstream programs for insufficient attention to phonetically regular words. As Chall concluded in her review of updated copyrights of the textbook programs she had critiqued 15 years earlier, the textbook publishers responded to her critique by increasing the phonics instruction in the teacher's manuals, but not substantially changing the phonics content of student texts.

Beck and Block (1979) confirmed Chall's observations in their analyses of the same generation of textbooks. Their analyses used a construct called "potential for accuracy" as a means of establishing the correlation between lessons in the teachers' edition and student books. This construct defines a word as decodable if its letter-sound elements have been taught in the teacher's guide. In applying this criterion to beginning reading texts from phonics and mainstream basal programs, the researchers concluded that 69-100% of the words in phonics programs had the potential to be accurately decoded, while 0-13% of the words had the potential

for accuracy in mainstream programs. The correlation conducted in this study between teachers' manuals and student texts addressed neither the number of lessons that particular groups of students required to assimilate a new phonics element nor the number of repetitions of words or spelling patterns that children required to recognize words independently.

Juel and Roper/Schneider's (1985) study is often cited as evidence for the potential for accuracy construct. Unlike the potential of accuracy model, however, this study did not examine the relationship between the content of instructional lessons and the phonetic regularity of the words in the texts. Instead, decodable words were defined as those words with predictable letter-sound patterns based on common phonics generalizations. They used this definition to study the reading acquisition of children who received the same phonics lessons, but who read from either a mainstream basal program or a phonics program. Juel and Roper/Schneider reported that the regularity of letter-sound relations in the words of a text and the number of times words were repeated influenced children's early word recognition.

Despite a lack of research that establishes the number of exposures to phonics content is required, the potential for accuracy criterion was the basis for selecting beginning reading programs in the most recent Texas textbook adoption (Texas Education Agency, 1997). Since coverage in a lesson was considered sufficient for student acquisition, programs complied with the mandate if 80% of the words in texts had words with letter-sound relations that had been covered in that lesson or in previous ones. Repetition of particular elements or of words was not tallied. Foorman, Francis, Davidson, Harm, and Griffin's (2002) analysis of three Texas adopted programs as well as of three others (including a prior copyright of a Texas adopted program) showed that programs varied considerably in their potential for accuracy rates. Further, often the criterion was achieved through holistically-taught words.

TExT Model

As the review of existing perspectives on text scaffolds suggested, there are many questions about the relationship of text to the processes and content of successful beginning reading. While the controlled text of the pre-primers and primers designed by Gray et al. (1956) and other teams of writers are gone, the mandates for decodable text have been juxtaposed with the interest of teachers in engaging literature. Thus, anthologies remain the core components in the Texas-adopted first-grade programs. Many of the texts in these anthologies are highly predictable, but also present phonetically regular words. Whether beginning readers attend to letter-sound correspondences carefully when they can rely on the predictable sentence or text structure is uncertain. What is needed is the systematic development of a model that considers factors emphasized in prior theory and research but that, in addition, has a sufficiently solid theoretical grounding that ensures that all critical factors are included.

The focus of the TExT model is on describing linguistic and cognitive proficiencies that children require to be successful with texts and also the proficiencies that texts foster in beginning readers. Specifically, the TExT model postulates two critical constructs in determining beginning readers' success with texts: (a) linguistic content (e.g., types of words) and (b) cognitive load (e.g., number of different words and number of repetitions per word).

Linguistic content. Linguistic content refers to the knowledge about oral and written language that is required to recognize the words in particular texts. Differences in linguistic content are evident in two types of words: (a) phonetically regular words and (b) words that occur frequently and often contain irregular letter-sound correspondences. These two groups of words can be changed by adding morphemes (i.e., inflected endings and comparative suffixes). At the very beginning stages of reading, morphological changes other than plurals, possessives,

and inflected endings are infrequent. The degree to which multisyllabic words are present in beginning reading texts also require attention because of the role that these words can have in encouraging a visual word recognition strategy (Juel & Roper/Schneider, 1985).

At its core, written English is an alphabetic system where one or more letters represent specific sounds in spoken English. When children are taught about the most consistent and common of these letter-sound relations, beginning readers learn more quickly to recognize words (Snow, Burns & Griffin, 1998). Research findings support this conclusion but numerous questions remain about precisely how many of the relations between the 44 phonemes and the 26 letters of English need to be taught. Some phoneme-grapheme relations appear in only a small group of words. For example, most beginning reading programs emphasize a primary phoneme for each letter of the alphabet. Each alphabet letter receives the same treatment even though some, like the letter z associated with the phoneme /z/ occurs in only one word in the 1,000 most frequent words in written English (Carroll, Davies & Richman, 1971; Zeno, Ivens, Millard, & Duvvuri, 1995). Other phoneme-grapheme combinations appear with greater frequency in written English but are not addressed systematically in programs. The choices made by textbook publishers on the number of phoneme-grapheme relations that are addressed in the lessons recommended in the teachers' guides of three reading programs differ, as Foorman et al. (2002) have shown. The numbers ranged from a low of 55 in one program to a high of 125 in another program. To date, research has not been conducted to establish differences in children's learning from programs differing in their treatment of grapheme-phoneme relations. Neither have studies established the appropriate number of phoneme-grapheme relations that support growth at particular points in beginning readers' development.

The TExT model considers two measures of different but complementary information on

the phoneme-grapheme knowledge required to read a text. The first description of phoneme-grapheme knowledge summarizes the complexity of the vowel patterns in a text. American phonics instruction has moved typically from simple vowel patterns to complex vowel patterns in single-syllable words. Simple vowel patterns are those in which a single phoneme is represented by a single grapheme such as the Consonant-Vowel-Consonant (CVC) in *cat*. The degree to which texts exaggerate particular vowel patterns at particular points such as simple vowels represented by the five different phoneme-graphemes: *a, e, i, o, u,* is captured by this measure of vowel complexity.

The second aspect of phoneme-grapheme relationships that the TExT model considers, is the degree to which highly common vowel and consonant patterns are repeated. Juel and Roper/Schneider (1985) attributed the success of the phonics basal program in promoting faster reading acquisition to what they termed a word family approach and what subsequently has been called a "rime" approach. Rimes consist of a vowel and any consonant (s) that follow it, while word families are the group of words that share a rime. Juel and Roper/Schneider's finding on a phonologically successful word recognition strategy when beginning readers read from texts with repetition of common rimes with different onsets or beginning consonants has been substantiated by Thompson, Cottrell and Fletcher-Finn (1996), in learning individual words.

The TExT model considers the "instantiations" or number of different onsets that appear with a rime. In the text Run, Run (Cassidy et al., 1987), the lines "Run, Nan. Run, Dad." have three short vowel rimes, none of which has more than one instantiation. The two lines in the text My pet is sick , "My pet dog is sick. Get the vet!" also have 3 rimes. In this case, there are 3 instantiations of one of the rimes (et) and two rimes with one instantiation each (*og, ick*).
For children to extend their use of these larger chunks of words, the occurrence of rimes with

different onsets (i.e., consonant(s) that precede the rime) may be more critical than seeing the rime in the same word repeatedly. Juel and Solso (1981) showed that exposure to words that share a rime such as *man, can, van,* and *tan*, rather than the repetition of a single word such as *ran* leads to application to new words with the same pattern.

While the knowledge of consistent letter-sound correspondences constitutes the primary linguistic content required to become a proficient reader of English, application of this knowledge needs to be accompanied with a set for diversity. The most-frequent words in written English are full of irregular correspondences, some of which occur only in these words. Consider the vowel patterns in the five most-frequent words in written English: *the, of, and, to*, and *a*. The vowel patterns in four of these five words are irregular, and occur only in these words, or in a handful of other words. Since these five words account for 17 out of every 100 words of text (Zeno et al., 1995) and an additional 20 words (6 with irregular vowel patterns) account for another 16 out of every 100 words of text (Zeno et al., 1995), children will be confronted early and frequently with the need for flexible application of linguistic content.

The number of syllables in words is a third aspect of linguistic content that would be expected to influence beginning readers' recognition of words. Multisyllabic words can be challenging for beginning readers who are in the process of developing decoding strategies. Juel and Roper/Schneider (1985) reported that the number of syllables accounted for more variance and decodability of monosyllabic words less variance in the mainstream basal which had a high percentage of multisyllabic words than in the phonics basal which had a low percentage of multisyllabic words. Since the 1970s when the texts studied by Juel and Roper/Schneider were published, the number of multisyllabic words has increased in texts for beginning readers

(Foorman et al., 2002). To acquire fluent word recognition strategies early on, texts with lower percentages of multisyllabic words are deemed preferable to those with higher percentages.

Cognitive load has to do with the amount of new linguistic information to which beginning readers can attend while continuing to understand the text's message. Repetition of at least some core linguistic content has been assumed in teaching children to read. The research that underlay the behaviorist model claimed that children required 35–45 repetitions of a word in order to recognize it (Gates, 1930; Gates & Russell, 1938a, 1938b). The words in these studies were exclusively high-frequency words, such as, *the, then, there,* and *they*.

Beginning in the 1970s, cognitive science perspectives were applied to models of word recognition and beginning readers' processing of text (Ehri, 1991; Stanovich, 2000). Applying these perspectives, beginning readers have been found to process and retain words in memory differently as a function of the number of words that are already recognized, the characteristics of words (lexical as well as letter-sound constituents), and the ratio of known to unknown words in a text (see Hiebert & Martin, 2001, for review).

Research on these cognitive perspectives has been largely confined to experimental texts (Reitsma, 1983). The application of cognitive processing models with the typical texts of instruction has been infrequent, and the repetition of words in beginning texts has been a non-issue in recent discussions of appropriate texts for beginning readers. Even though decodable texts have replaced the predictable and literary texts that followed the mandate for authentic, not controlled, texts, the repetition (or lack thereof) of decodable or high-frequency words has not figured into state textbook mandates or into publishers' design of programs.

The current context is described by Foorman et al. (2002) whose initial aim of describing text decodability in current programs was redirected to attend to the number of different words

and the lack of repetition. According to Foorman et al.'s (2002) review of six textbook programs, 70% of the unique words in a six-week instructional unit consisted of non-decodable words that occurred a single time. Even in the program with the fewest number of unique words, 50% of the words appeared only once in each six-week instructional block across the year. Since, as Hiebert, Martin, and Menon (in press) have reported, the number of unique words in the first 10 texts of grade one, regardless of philosophy (literature or phonics) averages 20 words per 100, beginning readers see many words, only a few of which are repeated. Beginning readers may receive practice with words in other contexts (e.g., word cards, word lists) but, when encountering texts, they are expected to recognize many words in their first encounter with them.

The guidelines from the existing literature on the number of different words within a text appropriate for beginning readers and the number of times different types of words should be repeated are almost nonexistent. Share's (1999) conclusion that children require as few as four to six exposures to respond rapidly to words was based on a study of average second graders. Reitsma's (1983) conclusion of four to six repetitions of words as optimal similarly came from normally progressing children in the second half of first grade, a time when normally progressing readers have attained automaticity with a core group of words. To date, an appropriate level of word repetition has not been established for children at the very beginning stages of reading.

Repetition of words and the number of unique words relative to total words are means of understanding the cognitive demands for beginning readers. Two other features of the texts that are commonly used in classrooms that have been described as lessening the cognitive demands on children and increasing their cognitive engagement are illustrations, and sentence and text structure. Numerous questions about the role of illustrations and sentence and text structure remain that require empirical investigation. In this study, the aim was to describe the potential of

the illustrations and sentence and text structures in supporting children's recognition of words. The parameters of the study did not permit an investigation as to the degree to which these features alleviate or exacerbate cognitive load for beginning readers.

The Current Study

This study is an investigation of the effects on young children's beginning reading as a result of participating with little books that have been sequenced according to an underlying curriculum of linguistic and cognitive knowledge or with the anthologies adopted by their school district. The sorting of little books occurred primarily in relation to a curriculum of linguistic content. In that cognitive load has not been considered in the writing of texts in recent decades, it was not possible to sequence existing books according to this variable to the degree that might be optimal. Students in the intervention condition read from little books organized along this underlying curriculum, as a function of their reading proficiencies. Teachers in the comparison classrooms continued to use the district-adopted anthology textbook program. During the year of the project, this textbook program was the most widely purchased by American schools (Education Market Research, 2000). Furthermore, this program presents research evidence that it has been effective in an inner-city school system with similar demographics to the district in which the target school was located (Houghton Mifflin, 2002).

We addressed three questions in this study: (a) What are the differences between the word-level characteristics of a little book curriculum sequenced according to linguistic knowledge and, to the extent possible, cognitive load and those of a prominent basal anthology series? (b) Is the re-sequenced little book curriculum effective in scaffolding independent word-solving skills relative to the progress made in a basal anthology curriculum? and (c) How effective are the two programs in supporting children's attainment of particular standards?

Method

Site and Participants

The participants in this study were 75 first-graders (100% African American; 50% on reduced/free lunch program) and their teachers, all from four first-grade classrooms of a K-5 charter school in inner city Detroit. The school was selected for its location in a high poverty district and, similar to other schools in the district, its use of a prominent literature anthology program, Houghton Mifflin's Invitations to Literacy (Cooper et al., 1998). This textbook program includes little books and several other little book programs were available in the classrooms and the school's resource room. Observations made during the semester and the year prior to the study indicated that little use was made of these little books. Teachers relied on the anthology of the literature-based program for reading lessons.

The four first-grade classrooms in the school were grouped into two intervention and two comparison classrooms. Teacher experience was the primary criterion for this assignment, with the classes of a first-year teacher, and a teacher with three or more years of teaching experience assigned to each condition. Pre-test scores of students in each condition were analyzed to determine that the two groups had approximately similar levels of reading proficiency at the initiation of the intervention. Children in the intervention classrooms (n=39) were provided with little books matched to their reading levels over a 15-week period (January-May). Children in the comparison classrooms (n=36) continued to read texts from the anthology of the literature-based reading program.

The Curriculum and Texts of the Intervention

As has already been stated, Houghton Mifflin's Invitations to Literacy (Cooper et al., 1998) was the school's adopted textbook program. The texts for the intervention needed to come

from existing sources since financial resources were not available to design, implement, and publish a set of original books based on the theoretical framework. A little book program--*Ready Readers* (Hiebert, Juel, & Englebretson, 1997) — was chosen because it claims to balance phonics content and high-frequency words with the leveling criteria of Reading Recovery (Peterson, 1991): theme, picture-text match, predictability of text structure, and language style. Relative to other little book programs, including the prominent ones in the marketplace, such as the little books of Rigby Education (2000) and Wright Group (1996), this program attempted to focus on linguistic content. However, word-level elements of cognitive load had received short shrift relative to the leveling criteria of Reading Recovery.

Since the Ready Reader program had approximately three times as many books as were needed for the 15-week intervention period—300 in the program as compared to the approximately 125 books needed for the intervention—the books were reordered to provide the best possible scaffolds to young children's reading acquisition. The process of reordering the books was a two-phase process. While research is clear that focused linguistic content is critical to young children's reading success, research is less forthcoming on the curriculum or specific sequence of content that should best be followed. The first phase was to choose a curriculum of linguistic content. The second phase was to apply this curriculum to the available books and to select books that best fit the linguistic curriculum, while attending as much as possible to the cognitive load of the books.

Choosing a curriculum of linguistic content. A previous review had shown that many state frameworks and textbook programs fail to specify the specific linguistic content with which readers need to be facile at particular grade levels (Hiebert, 2002). That review concluded that, without clear guidance from educational agencies, tests are the ultimate determiners of students'

attainment of grade-level reading standards. Consequently, the critical linguistic content for interim points in first-grade was established from an analysis of five prominent first-grade assessments that provide at least three levels for first grade: (a) Developmental Reading Assessment (Beaver, 1997); (b) Gray Oral Reading Test (GORT-4) (Wiederholt & Bryant, 2001); (c) Basic Reading Inventory (Johns, 1997); (d) Qualitative Reading Inventory (QRI) (Leslie & Caldwell, 2001; and (e) Texas Primary Reading Inventory (TPRI) (Texas Education Agency, 2001). For the end-of-grade-one, four additional assessments that attend only to this level were used: (a) TerraNova California Achievement Test (CAT-6) (CTB/McGraw-Hill, 2002); (b) TerraNova Comprehensive Test of Basic Skills (CTBS) (CTB/McGraw-Hill, 1997); (c) Gates-MacGinitie Reading Test (MacGinitie, MacGinitie, Maria, & Dreyer, 1998); and (d) Stanford Achievement Test (Sat-9) (Harcourt Brace, 1996).

The content for three milestones of first-grade, each representing the end of a trimester, was established through analyses of the preprimer, primer, and first-grade assessments. The criterion for linguistic content at a milestone (such as the preprimer level) was a curriculum that accounted for 85% of the unique words across the available assessments. For the preprimer level, facility with the 100 most frequent words and words with simple-vowel patterns was required. For the primer level, facility with the 200 most-frequent words was required and words with long-vowel patterns. For the grade-one level, the word recognition of the tests covered the 300 most-frequent words and monosyllabic words with complex and variant vowels.

The three levels of assessments capture end-points for phases of the first-grade curriculum: preprimer or end of the first trimester, primer or end of the second trimester, and first-grade or end of the third trimester. A curriculum supports beginning readers in achieving these milestones by ensuring that the knowledge represented by this milestone is covered in the

preceding materials or lessons. The development of the first milestone--reading at the preprimer level--represents a substantial, qualitative change in children's reading when they move from emergent literacy behaviors to conventional reading. In recognition of the substantial amount of growth required for the first milestone, the linguistic content associated with the preprimer milestone was broken down into three levels. Since the other two milestones represent less substantial changes, qualitatively, the linguistic content for the primer and first-grade levels was developed over two levels. The content for the seven levels is summarized in Table 1.

Insert Table 1 about here

Selecting texts for the little book condition. To identify the appropriate texts from the Ready Reader program that would support the designated linguistic content and cognitive load, the 300 texts in six levels of the Ready Reader program were analyzed for the features of linguistic content (high-frequency words, phonetically regular words) and cognitive load (unique words per 100, percentage of singletons). Since levels were sufficiently differentiated and several texts were to be read each week, the anticipation was that children would move through several levels during the intervention. The goal was to identify approximately 20 texts for each of the seven levels.

Texts were first sorted for their match on linguistic content. All of the 300 texts were analyzed with the TExT Analyzer (Hiebert & Martin, 2002). The TExT Analyzer is a HyperCard program that provides summary data on these features of texts: the total number of words, the number of unique words, and the number of repetitions of unique words. The TExT Analyzer

provides the ranking of the frequency of a word according to Carroll et al. (1971) and confirmed, through the first 300 words, with the more recent Zeno et al. (1995) list.

In establishing linguistic content of a text, high-frequency words in the designated curriculum were established first. Since many of the high-frequency words have irregular vowel patterns, particularly in the first 100, high-frequency words were not included in the analysis of vowel patterns of unique words. The TExT Analyzer describes the vowel patterns of unique words on an 8-point scale: (a) simple long vowel (e.g., go), (b) simple short vowel without blends or digraphs (e.g., at, cat), (c) simple short vowel with blends or digraphs (e.g., chat, bath), (d) long vowel represented by two graphemes, without blends or digraphs (e.g., meet, ride), (e) long vowel represented by two graphemes, with blends or digraphs (shine, teeth), (f) r-controlled vowels (e.g., car), (g) vowel diphthongs (e.g., oil) and variant vowels (bread), and (h) multisyllabic words (e.g., geranium). For purposes of analyses, the first three categories are presented together as simple vowel patterns, the next two categories as long vowel patterns, and categories (f) and (g) as complex vowel patterns. Multisyllabic words have their own classification. The TExT Analyzer also lists all of the words that share a rime. For example, the summary data for a text such as Little Bear indicates that there are 6 rimes with an average of 2.2 instantiations.

The initial aim was to have the same level of unique words meeting the core linguistic content as on the assessments—85%. However, analyses quickly indicated that this level could not be attained. Two criteria were then established. The first was to achieve a minimum level of 65% of the unique words representing the core curriculum. Another objective was for the average number of unique words that were multisyllabic and that occurred a single time in a text to not exceed 10%. The manner in which the texts of the little book program matched the core

linguistic curriculum is summarized in Table 1. The characteristics of the texts from the literature anthology are also summarized in Table 1. In that the comparison teachers were following the literature anthology sequentially from the beginning to the end of the school year, the 15-week period of the study coincided with the primer (4-5) and first-grade (6-7) levels.

As can be seen in Table 1, the criterion of 65% of the words in either the high-frequency or phonetically regular word curriculum was achieved for all but one level of the reclustered little books—Level 3. Level 3 had been established as the end of the preprimer level. While the texts contained many simple-vowel words (the curriculum of the preprimer level), the Level 3 texts had increased substantially in length from the two prior preprimer levels and had begun to integrate long-vowel words. Since many more vowel patterns in monosyllabic words were part of the core vocabulary at the primer and first-grade levels, percentages of words that fit the target curriculum would be expected to be quite high at these points. This pattern was true for the texts from the little books but less so for the texts from the literature anthologies.

The repetition of rimes with different onsets (i.e., instantiations of rimes) also determined the selection of texts from the little book program. As the data in Table 2 show, 36% of the unique words across the seven levels were members of word families with at least two different onsets (the criterion for inclusion in this category). The objective was to have at least one-third of the unique words instantiate common rimes at particular levels of the little books. By contrast, 21% of the unique words at the three levels of the anthology program met this criterion.

Once texts had been sorted for linguistic content, they were examined for cognitive load. Cognitive load features of the 125 little books that were identified for use in the study and the texts from the literature anthology are presented in Table 2.

Insert Table 2 about here

Cognitive load features were used to sort texts in the little book program, within levels with similar linguistic content (such as Levels 1 through 3). The first cognitive load feature considered, was the number of new unique words per text. Since texts with fewer unique words had fewer total words, the number of total words also increased through the levels. Since a core group of high-frequency words accounts for a high portion of the total words of a text, the more words in a text means that there are likely fewer unique words. Hence, long texts tend to have higher levels of repetitions. Typically, these repetitions are accounted for by numerous repetitions for the top 10 to 25 high-frequency words. At the earlier levels, a minimum of 4 repetitions per word was set as the baseline.

Sentence-text structure and picture-text match also needed to be considered as easing or increasing cognitive load, since the little book program had been written to emphasize these features. To describe the support provided by text-sentence structures, a scheme with a proven record of reliability was used (Hoffman et al., 1994). As described by Hoffman et al., two raters classified each text for the presence ("yes") or absence ("no") of nine features of predictability--repeated pattern, familiar concepts, cues from text for vocabulary, rhyme, rhythm, alliteration, cumulative pattern, familiar song/story, and familiar sequence. These ratings were used to establish a holistic score from 5 (highly predictable text where multiple features are present) to 1 (no evidence of predictable characteristics). Inter-rater agreement in choice of final category was high--92%. As can be seen in Table 2, texts in the little book and literature anthology programs had a similar rating for predictability at the primer level where both attained an average rating of

3. This rating signifies predictability through prominent use of one or two features such as a repeated pattern of sentences or phrases or a cumulative sequence in the Hoffman et al. scale. While the texts of the little program steadily showed a decrease in predictable features through the first-grade levels, the texts of the anthology program vacillated. Some of the texts in the first-grade levels (6 & 7) of the anthology program were highly predictable, while others continued to be highly predictable even at the end of the program.

While a picture-text progression has been emphasized in the Reading Recovery (Peterson, 1991) and guided reading schemes (Fountas & Pinnell, 1999), an explicit set of categories with established reliability indices has not been reported in the research literature. A scheme that quantifies the role of the picture in supporting identification of words was developed to describe the texts used in this study. The paradigm used for the scheme was to establish the number of words that adults wrote down in a 30-second exposure to a two-page spread of a text's illustrations. The exact matches between words generated and the words on the page (or "hits") were established. The picture-text match was the number of hits divided by the number of words on the page. For example, if raters identified the word "dog" from the illustrations associated with the sentence, "My pet dog is sick.", the picture-text match would be 20%.

Two experienced teachers who were unfamiliar with the texts responded to the illustrations on the middle page or pages (in cases of a two-page spread) of each of the 125 little books and the 12 literature texts used in the two conditions. Texts from the seven levels of the little book texts and the three levels represented by the literature texts were presented in a randomized order to the raters over five sessions. The raters were given 30 seconds per page to list all of the words that were elicited from the illustration.

An average of hits by the two raters was established for each of the 125 texts of the little book intervention and for the 12 texts in the anthology condition. The average percentage for picture-text match across the texts at a level is included in Table 2. As can be seen, the percentage of words that could be named or identified from pictures was approximately one of four or five words in the preprimer levels for the texts of the little book programs, about one of six to eight words at the primer levels, and about one of 20 to 25 words in the first-grade texts in the little book program. The support provided by pictures was fairly consistent and remained at a moderate level—one of every eight words—in the texts of the anthology program.

Documenting texts read in the two conditions. One of the two investigators observed in each classroom three times weekly during the duration of the study. As part of the observations, the investigator noted the texts that students read. On the days that the investigator did not visit, teachers kept a log of the texts read by their students. The logs of teachers and investigators were used to establish the enacted text-based curriculum—that is, the features of texts that students experienced over the study.

The enacted text-based curriculum for the comparison students was established in the following manner. Both teachers in the comparison classrooms used the district-selected textbook program, Houghton Mifflin (Cooper et al., 1998). The classes proceeded at slightly different paces through the anthology. Ms. Sumter, the first comparison teacher, used the textbook anthology for whole-class instruction. The second comparison teacher, Ms. Jarrod, had three reading groups that were at different places in the anthology when the study started in mid-year. By the end of the study, the high group in Ms. Jarrod's class and all of the students in Ms. Sumter's class had read the same 12 selections in the anthology. The average group in Ms. Jarrod's class had read 11 selections and the low group had read 10. When these 12, 11, and 10

selections were divided across the three time periods, the text features were not substantially different for the low, average, and high groups in Ms. Jarrod's classroom and for Ms. Sumter's class. To ensure clarity of presentation, the enacted text-based curriculum presented in this paper for the comparison students consisted of the 12 selections that were read by all of one class and 35% of the second class. For 65% of the second class, total numbers of words read were somewhat lower. The types of words read, however, were similar.

Students were assigned little books in the intervention classrooms according to their achievement level on the initial assessment. Each class had three groups but the initial achievement levels were different in the two classrooms. The groups in Ms. Riley's classroom are part of the very low, low, and high groups in the subsequent presentation of data, while the groups in Ms. Lindbergh's classroom were low, average, and high. In that the little books came from the same designated curriculum, the low groups in the two classrooms read roughly the same texts during the same time periods. The same was true for the high groups in the two classrooms. In the cases of the very low and the average groups, the data represent a distinct group in each of the two classrooms.

Instructional context

Intervention. The content of the intervention centered on the change from the district's adopted textbooks to the curriculum-based little books. Intervention teachers were asked to commit to providing their students from three to five little books consistently for 15 weeks. In addition, teachers were requested to ensure that children read books at their reading levels. They were provided with multiple copies of the 125 little books that had been leveled according to the designated curriculum. They were also apprised of the investigators' evaluations of students, based on the initial assessments, relative to the seven levels at the beginning of the 15-week

intervention period. In weekly visits to both groups of classrooms, one of the investigators discussed students' performances relative to the instructional texts, with the teachers.

The two teachers were given flexibility in how they organized their classrooms for the little book sessions and in the instruction that they provided around the little books. Instruction in the two intervention classrooms had several important similarities and differences. Each teacher used the little books in 25 to 30 minute time blocks. In Ms. Lindbergh's classroom, the little book sessions occurred daily. A daily session began with students locating the little books that were designated for their group for that week. They spent the daily session reading one new book and one or two books that they had read previously either independently or with partners. As in the comparison classrooms where taped versions of the texts were available, taped versions of the little books were made available to intervention teachers. Children of different levels spent at least one or two times weekly listening to one or more of the designated books for the week. Ms. Lindbergh also met with each of the three little book groups once a week. During these sessions, Ms. Lindbergh discussed strategies as students encountered unknown words in their reading of a particular little book. Ms. Lindbergh continued to provide whole-class phonics lessons that were not necessarily tied to the designated curriculum in instructional sessions that preceded the little book reading. Other texts, typically big books, were the focus of these phonics lessons. When she was not reading with one of the three groups, she listened to children read individually from the little books.

In Ms. Riley's intervention classroom, students read a different little book on each of the three days of the week. The specific book for each group was distributed at the beginning of a session and students read with partners in their group, independently, or following along with a tape-recorded version. Ms. Riley spent the entire session with a different group on each of the

three days. These small-group sessions were devoted to round-robin reading with few comments about strategies or content. On the fourth day of the week, all of the students in the class reviewed their books from that week, or engaged in word-card activities based on the little books they had read that week.

Comparison. In the comparison classrooms, children read approximately one text from the anthology per week in lessons that averaged an hour per day. Ms. Jarrod allocated a different text weekly to students according to their membership in one of three achievement groups. The lowest group was given more time with a selection before being moved on to the next one, while the higher groups moved at a slightly faster pace through the selections. Ms. Sumter, the second comparison teacher, moved students through the texts on a whole class basis. On the first day of the week, she read the text to students as they followed along. The remaining sessions of the week were devoted to rereading the text along with a tape-recorder or with partners. In both comparison classrooms, part of the daily hour-long sessions also involved whole-class lessons on word-level skills. Difficult words from the anthology selections were selected as the spelling words of the week, which children were asked to memorize. Similar to the pattern in the intervention classrooms, weekly visits by the investigator to the two comparison classrooms was followed by a discussion with teachers as to students' performances on the instructional texts.

Assessments

Narrative texts and word lists from the QRI (Leslie & Caldwell, 2001) were used as the pre- and post-test assessments. The QRI begins with a text for the preprimer level, followed by texts for the primer and first-grade levels. Pilot testing with the previous cohort of first graders in the spring of the year had shown that the preprimer level was too difficult for a portion of the cohort. Consequently, a number of texts were pilot tested to identify two that could precede the

preprimer as an assessment. Of this set of texts, two were found to be easier than the preprimer text of the QRI. Further, these two texts elicited different performances from beginning readers. That is, the first text was easier than the second text and both were easier than the preprimer text of the QRI. The first text was entitled Monster Mop (Mark, 1997) and the second was Good Girl (Jacobs, 1997). These two texts were administered to students who were unable to attain 90% accuracy on the preprimer level passage of the QRI. Word-lists consisted of the 15 most-frequent words on each of the QRI word lists.

The narrative forms of the QRI were used for both the pretest and posttest. This choice was made because of the lack of equivalence of the QRI narrative passages in difficulty. Further, the time gap between the two administrations was sufficient so that rarely did the same child end up reading the same passage (due to greater proficiency by the end of first-grade).

Both the QRI word lists and passages yield scores that are ordered, but discontinuous. In order to create an equivalent scale, students were assigned a score for the final level that they read. If students read the first early literacy text, Monster Mop, with appropriate levels of fluency and accuracy but failed to attain criteria on the second early literacy text, they were assigned a score of 1. Successful performance on the second early literacy text, Good Girl, but not on the third text which was the preprimer text of the QRI, was scored as 2. The QRI levels were scored as 3 (preprimer), 4 (primer), 5 (first grade), 6 (second grade), 7 (third grade), and 8 (fourth grade). No student attained the criteria of accuracy, fluency, and comprehension beyond the fourth-grade text of the QRI. A similar procedure was used for scoring the word lists. With the word lists, however, the lowest score was 1 for the preprimer word list of the QRI, and the highest score was 6 for the fourth-grade word lists.

Formation of Performance Groups

Pretest scores were used to assign students in each condition to four reading levels: very low, low, average, and high. Although reading on word lists and passages was highly correlated (r=+.80) at the time of the pre-tests, children were reading a little higher on the word-lists than on the passages.

Group formation also involved teachers' evaluations of student reading performances. Teachers were asked to evaluate their students as low, average, or high. Teachers evaluated average readers as those reading at primer level on word-lists and at preprimer text level in December. The remaining groups were established in relation to the average group. Because of the range of performances below the average level indicated by teachers, low and very low groups were formed. The low group consisted of children who were reading Passage 2, and the very low group students were reading at or below Passage 1. The high group also had substantial variation. For class organizational purposes, however, all students who were reading at or above a primer level text were assigned to the high group.

Results

Three questions were addressed in this study: (1) What are the differences between the word-level characteristics of a little book curriculum (re-sequenced according to the theoretical criteria of cognitive load and linguistic content) and a prominent basal anthology series in the tasks they pose to readers? (2) Is the re-sequenced little book curriculum effective in scaffolding independent word-solving skills relative to the progress made in a basal anthology curriculum? Are the effects of the two curricula different for children at different reading levels? and (3) Are there differences in the attainment of grade-level benchmarks between the two groups? The results are presented in terms of each of these questions.

Comparisons between Curricula

Post-hoc comparisons were made between the weekly text-based curricula for the four intervention groups (very low, low, average and high), and the comparison group. Data on text features were clustered into three five-week segments, representing the beginning, middle, and final phases of the 15-week intervention. The data on linguistic content and for cognitive load are presented in Table 3.

_____

Insert Table 3 about here

_____

Linguistic content. When considered relative to all of the unique words in an instructional unit, the percentage of multisyllabic words within a corpus indicates the difficulty of linguistic content of texts for beginning readers. As the data on multisyllabic words in Table 3 demonstrates, the enacted little book curriculum showed a progression in the difficulty level of the words introduced. In contrast, an average of 32% of the unique words in the anthology texts consisted of multisyllabic words from the first through the last time period. During the latter portion of the intervention, the average and high groups in the little book condition read from texts with levels of multisyllabic words that were equivalent to those of the anthology group. However, the percentages of multisyllabic words differed most for the intervention's very low and low groups: 12-15% of the words in the very low group curriculum and 11-20% in the low group curriculum relative to the 32% average of the anthology group.

While high percentages of multisyllabic words are likely to pose challenges for beginning readers, texts with high percentages of unique words that have rimes with the target vowel patterns permit beginning readers to apply phonics knowledge. Table 3 summarizes the

percentage of text derived from multiply instantiated rimes, that is, rimes with more than one onset. The anthology selections introduced children to fewer rimes with more than one onset than the four little book curricula. Further, these rimes were repeated fewer times in the weekly anthology curriculum (3.5-5 repetitions per rime), as compared to all levels of the little book curricula (7-10 repetitions per rime). Therefore, a smaller percentage of the text was made up of these multiply instantiated rimes in the anthology selections (7-16%), than in the little book curriculum (21-28%).

The analysis of the presence of the 100 most frequent words in written English (Carroll et al., 1971) revealed that the overall proportion of high frequency words in texts did not differ across the different curricula on weekly basis, averaging approximately 40-50% of the total text. However, the approaches did differ in the number of unique or distinct high frequency words that children were expected to read each week. Except for the very low group, texts in all other intervention groups had more distinct high frequency words per week than the anthology group texts. That is, the intervention group children saw more examples of high frequency words than students in the anthology group.

Cognitive load. As can be seen in Table 3, the data on cognitive load indicate that children in all intervention groups (except for the very low group) read more words per week than children in the comparison group. These words were distributed over a greater number of texts in the intervention condition than in the anthology condition. On average, intervention group children read between 15-20 little books per five-week period, while students in the anthology condition read four texts during this time. Even though the little book texts were considerably shorter than those of the anthology in the first two periods, the provision of more

texts in the little book condition meant that little book students were exposed to more words over the course of the study.

Despite the difficulty of finding texts with consistent word-density ratios for particular periods of time within the Ready Reader program, data in Table 3 show that a higher percentage of words was repeated in the little books than in the anthology texts: between 85-90% of all words in the former relative to 65-70% of the words in the latter. In particular, fewer words appeared a single time in the little book curriculum than in the anthology curriculum: 20% as compared to 30-35%. While the little book program did not provide a built-in, consistent progression in word density ratios across time and groups, the program did provide a greater degree of word repetition during a specific week than anthology texts read during a comparable time period.

Group Performances as a Function of Condition

Students' performances on the QRI word lists and texts from the beginning to the end of the intervention period were used to establish the effectiveness of the two conditions. Analyses of Covariance (ANCOVA) were conducted for each of the measures (word lists and text reading) to establish differences in mean post-test scores, using the pre-test scores as the covariates. The ANCOVAs were conducted using a nested design -- 2 X (4) X 3 [Text-Type X (Teachers/Classrooms Nested within Text-Type) X Reading Level]. The very low and low reading groups were combined for these analyses, and identified in terms of a broader category of struggling readers. Table 4 presents means on the two tasks by text condition, including adjusted means for the intervention and comparison groups that acknowledge the comparison group's somewhat higher (although not statistically different) pre-test performances on both the word and text measures.

Insert Table 4 about here

The main effect for treatment was significant at the $p < .01$ level for word reading [F (1,4) = 16.6] and text reading [F (1,4) = 10.6]. An examination of the means in Table 4 indicates that the students in the little book condition had higher means on both the word list and text measures than students in the anthology condition during the post-test assessments. The intervention group improved by 2.8 text levels as a result of the intervention, while the comparison group's improvement was 1.8 text levels during the same period of time. At the end of the 15 weeks, intervention students were reading, on average, texts of level 6 (second-grade), while comparison students were reading texts of level 5 (first-grade).

The main effect for reading level was not significant for either the text [F (2,4), = 2.56] or word reading tasks [F (2, 4) = 3.01], although it approached significance on the latter (.06). In considering the average gains on the word list made by the three achievement groups across the two text conditions, the greatest gains were made by the high-achieving students (an average of 1.9), while the average achievers with a gain of 1.4 made the smallest gains. The interaction for text condition by achievement level for both the word and the text tasks was not statistically significant. While achievement groups in the anthology condition did not make gains as large as those of students in the little book condition, students of the same achievement group made comparable gains relative to other achievement groups of their condition.

The plot lines in Figures 1 and 2 afford an opportunity to make comparisons across reading levels – that is, how did the different groups perform relative to each other?

Insert Figures 1 & 2 about here

An examination of Figures 1 and 2 reveals that the trajectory of progress was much steeper for all levels of intervention group readers on both the word lists and the text-reading tasks. The intervention group low readers were reading at similar levels to the comparison group average readers on both tasks by the time of the post-test. On the text reading task, the average readers in the intervention group had mean scores that were very close to the mean scores of high readers of the comparison group. These results indicate that each level of intervention readers was performing more similarly to the higher level of the comparison group readers, than to the group that they had started out with in December.

Teacher/classroom effect was examined by nesting it as an independent variable within the text-type. Effects for both word reading ($F(2,4) = 1.02$) and for text reading ($F(2,4) = 6.23$) were statistically insignificant, indicating that the results did not vary statistically within each treatment or text-type by teacher (or classroom).

Absolute Levels of Performance

While the analyses of mean differences indicate that students made higher gains in reading words and texts in the little book than the anthology condition, establishing the number of students who have attained particular benchmarks at the end of grade one is also critical. One such benchmark is proficient reading at the primer level on an informal reading inventory. Based on their text reading, students were placed into three groups: below-first grade (performances below primer), first-grade (primer and first-grade texts), and above-first grade

(second-grade and above). Percentages of students falling into these three groups according to the two conditions are provided in Table 5.

_____

Insert Table 5 about here

_____

For text reading, a chi-square analysis showed a significant difference (Chi-square (2, 75)=6.25, $p<.05$). A third of the students who read from anthologies were reading below the primer level in May. In comparison, 10% of the little book group was reading at this level. Further, two-thirds of the little book students were reading above the first group level. Ten percent fewer of the anthology group was reading at this level.

Discussion

The results of this study show that a moderate amount of scaffolding of texts can make a significant difference for first graders. Whether of initially low, average, or high reading achievement, students in the intervention group read at one level of text higher than the students in the anthology group by the end of grade one. After 15 weeks of reading from the little books, children attained an average of second-grade reading, while their peers who kept reading in the district's anthology read first-grade texts. Most of the students in the little book group were leaving first-grade able to be successful with the texts of the next grade, while their peers from the anthology group did not have this extra advantage.

Further, a third of the comparison group students had not attained the level of first-grade reading. By contrast, only 10% of the students in the little book group failed to attain the first-grade standard. To maintain high levels of reading proficiency, students will require instruction that builds on this foundation. However, if first-grade performance predicts fourth-grade

performance, a substantial portion of the students who read from little books, rather than the anthologies, have an advantage as they enter second grade.

As evidenced by the recent mandate of Texas for 80% decodable text, policy-makers and publishers have sought to identify precise prescriptions for beginning reading texts. The results of this study suggest that somewhat greater consistency in linguistic content. and somewhat less demanding cognitive demands support beginning reading acquisition. The results do not suggest, however, that beginning readers require texts where all words fit particular patterns or where each unique word is repeated a particular number of times. The little books were not written to a formula, nor were these texts ideal in their execution of particular features. Like the anthology texts, the little books had a significant percentage of words that were multisyllabic and multisyllabic words were the ones most likely to be repeated less than four times.

Even on those dimensions where the texts of the little book and the anthology programs differed, the differences were of degree, not of kind. The most substantial difference lay in the percentage of unique words accounted for by rimes with target vowel patterns. A quarter of the words in the little books consistently exemplified target vowel patterns. Percentages were low in the anthology texts, particularly at the beginning of the 15-week period. However, the percentage of words with target rimes in the little books did not achieve the critical mass of the phonics texts of an earlier era.

Similarly, the average repetitions per word of 6 for the little book program and 3 in the literature anthology program were far from the 35-40 repetitions recommended by Gates and Russell (1938a, 1938b). Despite this, the average word in the little book program was repeated twice as often as in the literature anthology program. A significant portion of the words in the latter program appeared a single time: one out of every five unique words was a single-

appearance, multisyllabic word. In the little book program, this figure was one of every 14 unique words—almost three times fewer single-appearing, multisyllabic words. Further, among the monosyllabic words, a higher percentage of the unique words in the little book program fit into families of words with common, consistent spelling-sound correspondences.

One area in which the texts of the two conditions differed substantially, was in the number of total words that students read. Except for the very low students, students in the little book intervention were given texts with substantially more words than students in the anthology classes. The average number of total words, around 250 words, for the latter classes was about a quarter of that read by the high students and about half that read by the average and low students in the little book intervention groups. Teachers in the anthology group had students repeatedly read the weekly text, including partner reading, reading along with tape recordings, and reading along with the teacher. The little books were to be read repeatedly, too. Because of the independent context in which students were often placed, this repeated reading did not occur consistently in the intervention classrooms. By repeatedly reading texts, the students in the anthology condition likely read as many total words as the students in the little book condition. The opportunity to reread texts has been emphasized as a technique for increased accuracy, fluency, and comprehension. The repeated reading literature has not systematically compared a model in which a handful of texts are being re-read, rather than many texts. Preliminary findings of a study where such a comparison is being examined suggest that, at least for beginning readers, reading a number of different texts may produce better results than reading a single text repeatedly (Stahl, 2002). More investigation is required but, with beginning readers who are reading texts with strong contextual clues (such as the picture-text match and text-sentence

structure), the technique of using multiple rather than single texts over a time period may be more beneficial.

The search for the precise prescription for texts for beginning readers is likely to be tedious. The factors involved in linguistic content and cognitive load, including variables not examined in this study such as imagery value of words (Sadoski & Paivio, 2001), are likely to be many and the permutations of variables even more. Guidelines for appropriate text, rather than the search for a precise prescription, will likely prove more fruitful than attempting to establish precise prescriptions. The findings of this study suggest two guidelines for texts for beginning readers: (a) texts should provide at least a moderate degree of consistency in linguistic content and (b) texts should provide at least a modicum of repetition of words and spelling patterns.

To some, these conclusions may fall into the realm of the "already known." However, the anthology program of the comparison condition is presented as the factor behind higher reading achievement in a research study conducted in Washington, D.C., a school system quite similar in demographics to that in the current study (Houghton-Mifflin, 2002). Further, the anthology curriculum defines the reading instruction that many children receive. Especially for students in less economically viable communities, the money provided by state funding agencies determines the materials that they receive. In purchasing textbooks in Texas (Texas Education Agency, 1997), the per dollar expenditure per student is for the anthology components. In California's recent adoption where decodable books needed to accompany the anthology, the anthology continues to be the central component in the teachers' manuals.

Texts in no way replace instruction. How teachers use texts is what matters. However, curricular materials play an important role in shaping/scaffolding teachers' instructional practices (Ball & Cohen, 1996), since they function as important tools that mediate instruction. Teachers

in the two comparison group classrooms used the texts in very different ways, indicating considerable teacher-discretion in the design of the enacted curriculum in the two classrooms. It is also important to consider that the reading programs that are available to beginning teachers may influence their understandings of what constitutes appropriate learning for that grade, as well as how to scaffold that learning.

On the other hand, the little books directed the attention of both teachers and children in the intervention group classrooms to critical word-level features of the texts. However, it is probable the form of the little books (usually marketed to teachers as "independent readers") led to their use in semi-independent formats in these classrooms. Teachers in both classrooms spent very little time per week reading the little books with the children, leaving them to read largely with their reading groups or with the read-along tapes. This suggests the need for further research to identify optimal instructional uses of these texts.

Much more empirical work with the underlying curriculum and book sorting is required before the strategies of this study are recommended to classroom teachers. For one, it is not clear at what point in reading development, the introduction of such a curriculum is most useful. All the children included in this sample had basic letter-sound knowledge, even though some children were not conventional readers at the start of the intervention. Further, the study did not examine the relevance of the curriculum past the end of first-grade. It may well be that students who reach the latter levels of this scheme should be reading from literature anthologies and from trade books, rather than little books, since the characteristics of texts in the two programs at Levels 6 and 7 did not differ substantially from one another. However, at the beginning levels, the findings of this study suggest that texts that have been crafted to support particular linguistic

content and that address the cognitive load of linguistic content can support higher levels of reading acquisition.

References

Anderson, R.C., Hiebert, E.H., Scott, J.A., & Wilkinson, I.A.G. (1985). *Becoming a Nation of Readers: The Report of the Commission on Reading*. Champaign, IL: The Center for the Study of Reading, National Institute of Education, National Academy of Education.

Ball, D. L., & Cohen, D. K. (1996). Reform by the book: What is--or might be--the role of curriculum materials in teacher learning and instructional reform? *Educational Researcher, 25(9*), 6-8.

Barr, R. (1972). The influence of instructional conditions on word recognition errors. *Reading Research Quarterly, 7*(3), 509-529.

Baumann, J. F., Hoffman, J. V., Moon, J., & Duffy-Hester, A. M. (1998). Where are teachers' voices in the phonics/whole language debate? Results from a survey of U. S. elementary teachers. *The Reading Teacher, 51 (8),* 636-650.

Beaver, J. (1997). Developmental Reading Assessment. Parsippany, NJ: Celebration Press.

Beck, I. L., & Block, K. K. (1979). An analysis of dimensions that affect the development of code-breaking ability in eight beginning reading programs. Pittsburgh: University of Pittsburgh, Learning Research and Development Center.

Brennan, A., Bridge, C., & Winograd, P. (1986). The effects of structural variation on children's recall of basal reader stories. *Reading Research Quarterly, 21,* 91-104.

Bridge, C. A., & Burton, B. (1982). Teaching sight vocabulary through patterned language materials. In J. A. Niles & L. A. Harris (Eds.), *New Inquiries in Reading Research and Instruction* (pp. 119-123). Washington, DC: National Reading Conference.

Bridge, C. A., Winograd, P. N., & Haley, D. (1983). Using predictable materials vs. preprimers to teach beginning sight words. *The Reading Teacher, 36*(9), 884-891.

California English/Language Arts Committee. (1987). *English-Language Arts Framework for California Public Schools (Kindergarten Through Grade Twelve)*. Sacramento: California Department of Education.

Carroll, J. B., Davies, P., & Richman, B. (1971). *Word frequency book*. Boston: Houghton Mifflin.

Cassidy, J., Roettger, D., & Wixson, K. K. (1987). *Join the circle*. New York: Scribner.

Chall, J. S. (1967/1983). *Learning to read: The great debate*. (3rd ed.). Fort Worth, TX: Harcourt Brace.

Chall, J. S., & Dale, E. (1995). *Readability revisited: The new Dale-Chall readability formula*. Cambridge, MA: Brookline Books.

Cooper, J. D., Pikulski, J.J., Au, K.H., Calderon, M., Comas, J.C., Lipson, M.Y., Mims, J.S., Page, S.E., Valencia, S.W., & Vogt, M. (1998). *Invitations to Literacy*. Boston, MA: Houghton Mifflin.

CTB/McGraw-Hill (2002). *TerraNova California Achievement Test (6th Ed.)*. Monterey, CA: Author.

CTB/McGraw-Hill (1997). *TerraNova Comprehensive Test of Basic Skills*. Monterey, CA: Author.

Education Market Research (2000). *Elementary reading market update: December 2000*. Retrieved April 14, 2003, from http://www.ed-market.com/r_c_archives/

Ehri, L.C. (1991). Development of the ability to read words. In R. Barr, M.L. Kamil, P.B. Mosenthal, & P.D. Pearson (Eds.), Handbook of research in reading (Vol. 2, pp. 383-417). New York: Longman.

Elson. W. H., & Gray, W. S. (1930). *Elson Basic Readers*. Chicago, IL: Scott Foresman.

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

Fountas, I., & Pinnell, G. S. (1999). *Matching books to readers: Using leveled books in guided reading, K–3*. New York: Heinemann.

Gates, A. I. (1930). *Interest and ability in reading*. New York: The Macmillan Company.

Gates, A. I., & Russell, D. H. (1938a). Types of materials, vocabulary burden word analysis, and other factors in beginning reading: I. *Elementary School Journal, 39*, 27-35.

Gates, A. I., & Russell, D. H. (1938b). Types of materials, vocabulary burden word analysis, and other factors in beginning reading: II. *Elementary School Journal, 39*, 119-128.

Goodman, K.S. (1968). The psycholinguistic nature of the reading process. In K.S. Goodman (Ed.), *The psycholinguistic nature of the reading process* (pp. 13-26). Detroit, MI: Wayne State University.

Gray, W. S., Monroe, M., Artley, A. S., Arbuthnot, A. H., & Gray, L. (1956). *The new basic readers: Curriculum foundation series*. Chicago: Scott, Foresman, & Company.

Harcourt Brace (1996). *Stanford Achievement Test (9th Ed.)*. Orlando, FL: Author.

Hiebert, E.H. (2002). Standards, assessment, and text difficulty. In A.E. Farstrup & S.J. Samuels (Eds.). *What research has to say about reading instruction* (3rd Ed., pp. 337-369) Newark, DE: International Reading Association.

Hiebert, E.H., Juel, C., & Englebretson, R. (1997). *Ready readers*. Parsippany, NJ: Modern Curriculum Press.

Hiebert, E.H, & Martin, L.A. (2001). The texts of beginning reading instruction. In S. Newman & D. Dickinson (Eds.), *Handbook of Research on Early Literacy* (pp. 361-376). New York: Guilford Press.

Hiebert, E. H., & Martin, L. A. (2002). *TExT (Task Elements by Task) software* (3rd ed.). Santa Cruz, CA: TextProject.

Hiebert, E.H., Martin, L.A., & Menon, S. (in press). Are there alternatives in reading textbooks? An examination of three beginning reading programs. *Reading Writing Quarterly*.

Hoffman, J. V., McCarthey, S. J., Abbott, J., Christian, C., Corman, L., Curry, C., Dressman, M., Elliot, B., Maherne, D., & Stahle, D. (1994). So what's new in the new basals? A focus on first grade. *Journal of Reading Behavior, 26,* 47-73.

Hoffman, J.V., Sailors, M., & Patterson, E.U. (2002). Decodable texts for beginning reading instruction: The year 2000 basals. *Journal of Literacy Research*, *34*, 269-298

Houghton Mifflin (2002). *Hough Mifflin reading program efficacy studies*. Boston, MA: Houghton Mifflin. Retrieved April 14, 2003, from http://www.eduplace.com/marketing/nc/research.html

Jacobs, S. (1997). *Good girl*. Parsippany, NJ: Modern Curriculum Press.

Johns, J.L. (1997). *Basic reading inventory (7th Ed.)*. Dubuque, IA: Kendall/Hunt Publishing Co.

Johnston, F. R. (2000). Word learning in predictable text. *Journal of Educational Psychology, 92*(2), 248-255.

Juel, C., Griffith, & Gough, P.B. (1986). Acquisition of literacy: A longitudinal study of children in first and second grade. *Journal of Educational Psychology, 78(4),* 243-55.

Juel, C., & Roper/Schneider, D. (1985). The influence of basal readers on first grade reading. *Reading Research Quarterly, 20 (2*), 134-152.

Juel, C. & Solso, R. L. (1981). The role of orthographic redundancy, versatility and spelling-sound correspondences in word identification. In M. L. Kamil (Ed.) *Directions in reading: Research and Instruction* (30th Yearbook of the National Reading Conference, pp. 74-92). Rochester, NY: National Reading Conference.

Leslie, L., & Caldwell, J. (2001). *Qualitative Reading Inventory - 3*. New York: Addison Wesley Longman, Inc.

Leu, D. J., Jr., DeGroff, L.J.C., & Simons, H. D. (1986). Predictable texts and interactive-compensatory hypotheses: Evaluating individual differences in reading ability, context use, and comprehension. *Journal of Educational Psychology, 78*(5), 347-352.

MacGinitie, W.H., MacGinitie, R.K., Maria, K., & Dreyer, L.G. (1998). *Gates-MacGinitie Reading Tests*. Itasca, IL: Riverside Publishing/Houghton Mifflin.

Mark, A. (1997). *Monster mop*. Parsippany, NJ: Modern Curriculum Press.

Peterson, B. L. (1991). Selecting Books for Beginning Readers. In D. E. DeFord, C. A. Lyons, & G. S. Pinnell (Eds.), *Bridges to Literacy: Learning from Reading Recovery* (pp. 119-147). Portsmouth, NH: Heinemann.

Reitsma, P. (1983). Word-specific knowledge in beginning reading. *Journal of Research in Reading, 6,* 41-55.

Rigby PM Plus Program (Starters, Red, Yellow, Blue, Green, Orange, Turquoise Levels). Barrington, IL: Rigby Education.

Sadoski, M., & Paivio, A. (2001). *Imagery and text: A dual coding theory of reading and writing*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Samuels, S. J. (1970). Effects of pictures on learning to read, comprehension and attitudes. *Review of Educational Research, 40,* 397-408.

Shanahan, T. (1984). Nature of the reading-writing relations: an exploratory multivariate analysis. *Journal of Educational Psychology, 76,* 466-477.

Share, D. L. (1999). Phonological recoding and orthographic learning: A direct test of the self-teaching hypothesis. *Journal of Experimental Child Psychology, 72,* 95-129.

Spache, G. (1974). The revised Spache readability formula. In G. Spache (Ed.), *Good reading for poor readers*. Champaign, IL: Garrard Publishing Company.

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

Stahl, S. R (November 2002). *Fluency: Instruction and assessment.* Presentation at the Focus on Fluency Forum, San Francisco, CA.

Stanovich, K. E. (2000). Progress in understanding reading: Scientific foundations and new frontiers. New York: Guilford Press.

Texas Education Agency. (1990). *Proclamation of the State Board of Education advertising for bids on textbooks*. Austin, TX: Author.

Texas Education Agency. (1997). *Proclamation of the State Board of Education advertising for bids on textbooks*. Austin, TX: Author.

Texas Education Agency (2001). *Texas Primary Reading Inventory*. Austin, TX: Author.

Thompson, G. B. Cottrell, D. S., & Fletcher-Flinn, C. M. (1996). Sublexical orthographic-phonological relations early in the acquisition of reading: The knowledge sources account. *Journal of Experimental Child Psychology, 62,* 190-222.

Thorndike, E. L. (1903). *Educational psychology*. New York: Lemcke & Buechner.

Thorndike, E. L. (1921). *The teacher's word book*. New York: Columbia University Press.

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

Wright Group (1996). Sunshine Reading Program. Bothell, WA: Wright Group/McGraw-Hill. Rigby Education (2000).

Zeno, S.M., Ivens, S.H., Millard, R.T., & Duvvuri, R. (1995). *The educator's word frequency guide*. Brewster, NY: TASA.

Table 1. Characteristics of Levels of Little Books and of Anthologies: Linguistic Content

Pro-gram	Level	High-Frequency Curriculum (%)			Phonically regular curriculum (%)			Target Lin-guistic Con-tent (%)	Multi-sylla-bic (%)	Instantiation of Rimes with Target Phonics Content		
		100	200	300	Sim-ple	Long	Com-plex			Rimes (#)	Instan-tia-tions (#)	Unique Words (%)
LITTLE BOOKS												
	1	25	-	-	44	-	-	69	6	5	4.0	32
	2	32	-	-	35	-	-	67	7	10	3.2	23
	3 (Pre-primer)	23	-	-	38	-	-	61	13	32	3.0	34
	4	17	14	-	31	13	-	75	16	60	3.2	43
	5 (Pri-mer)	15	11	-	28	17		71	14	76	3.6	35
	6	16	13	9	20	16	10	84	17	115	3.2	41
	7 (First-Grade)	11	9	7	20	15	13	75	25	100	3.4	41
ANTHOLOGY												
	4-5 (Primer)	23	13	-	15	11	-	62	29	15	2.5	17
	6	22	14	10	14	8	6	74	25	28	2.5	22
	7 (First-Grade)	20	16	8	14	9	4	71	26	36	2.7	25

Table 2. Characteristics of Levels of Little Books and of Anthologies: Cognitive Load

Pro-gram	Level	New Unique Words per text (#)	Total Words per text (#)	New Unique Words per 100 words (#)	Repetitions per word (M)	Text & Sentence Structure (Rating)	Picture Support (% of total words)	MS Singletons (%)
LITTLE BOOKS								
	1	6	23	26	4	5	28	6
	2	12	43	28	4	5	24	6
	3 (Pre-Primer)	16	78	21	5	4.3	21	5
	4	19	110	17	6	3.1	18	9
	5 (Pri-mer)	25	155	16	6	2.6	13	8
	6	39	229	17	6	2.3	5	8
	7 (First-Grade)	43	386	11	9	1	7	9
ANTHOLOGY								
	4-5 (Pri-mer)	66	228	29	3	3.2	19	21
	6	81	243	33	3	1.5	11	21
	7 (First-Grade)	96	325	29	3	1.9	15	19

Table 3. Enacted Curriculum: Text Features for Intervention and Comparison Groups at Three Time Periods

		Words per week (#)	Repeated words per week (%)	Multi-syllabic (% of total words)	Words derived from multiply instantiated rimes (% of total words)	Unique high-frequency words (#)
VERY LOW LITTLE BOOKS						
	Beginning	161	84	12	23	19
	Middle	255	85	16	21	27
	End	361	89	12	25	40
LOW LITTLE BOOKS						
	Beginning	267	84	11	26	26
	Middle	520	87	18	27	44
	End	653	86	20	27	58
AVERAGE LITTLE BOOKS						
	Beginning	327	85	14	26	34
	Middle	707	87	20	27	60
	End	1048	89	27	25	69
HIGH LITTLE BOOKS						
	Beginning	532	85	16	28	50
	Middle	1060	89	26	26	70
	End	1692	90	36	26	85
ANTHOLOGY						
	Beginning	213	70	32	7	26
	Middle	243	69	34	9	39
	End	307	65	29	16	39

Table 4. Means (and Standard Deviations) for Pre-Test and Post-Test Scores by Reading Level and Textual Condition

Groups	Word Lists		Passage Reading	
	Pre-test	Post-test	Pre-test	Post-test
Intervention group				
Low (n=16)	1.2 (0.4)	3.1 (0.4)	1.7 (1.0)	4.9 (0.8)
Average (n=10)	2.0 (0.0)	3.9 (0.6)	3.0 (0.0)	6.0 (0.5)
High (n=13)	2.9 (0.6)	5.5 (0.8)	4.9 (1.1)	7.1 (1.0)
Total (n=39)	2.0 (0.8)	4.1 (1.2)	3.1 (1.6)	5.9 (1.2)
Adjusted Mean		4.2		6.0
Comparison group				
Low (n=15)	1.3 (0.5)	2.6 (0.6)	1.7 (1.0)	3.9 (1.5)
Average (n=7)	2.0 (0.0)	3.1 (0.7)	3.0 (0.0)	5.3 (1.3)
High (n=14)	3.7 (1.1)	4.9 (1.6)	5.4 (1.3)	6.6 (1.2)
Total (n=36)	2.4 (1.4)	3.6 (1.5)	3.4 (2.0)	5.2 (1.8)
Adjusted Mean		3.4		5.2

Table 5. Percentages of Children Reading at Different Levels on Word Lists and Passage Reading Tasks

Tasks	Below First-grade	At First-grade	Above First-grade
Word Lists			
PRE-TESTING			
Intervention Group	74.4	23.1	2.6
Comparison Group	63.9	19.4	16.7
POST-TESTING			
Intervention Group	2.6	38.5	59
Comparison Group	16.7	47.2	36.1
Passage Reading			
PRE-TESTING			
Intervention Group	84.6	7.7	7.7
Comparison Group 75		5.6	19.4
POST-TESTING			
Intervention Group	10.3	25.6	64.1
Comparison Group	36.1	8.3	55.6

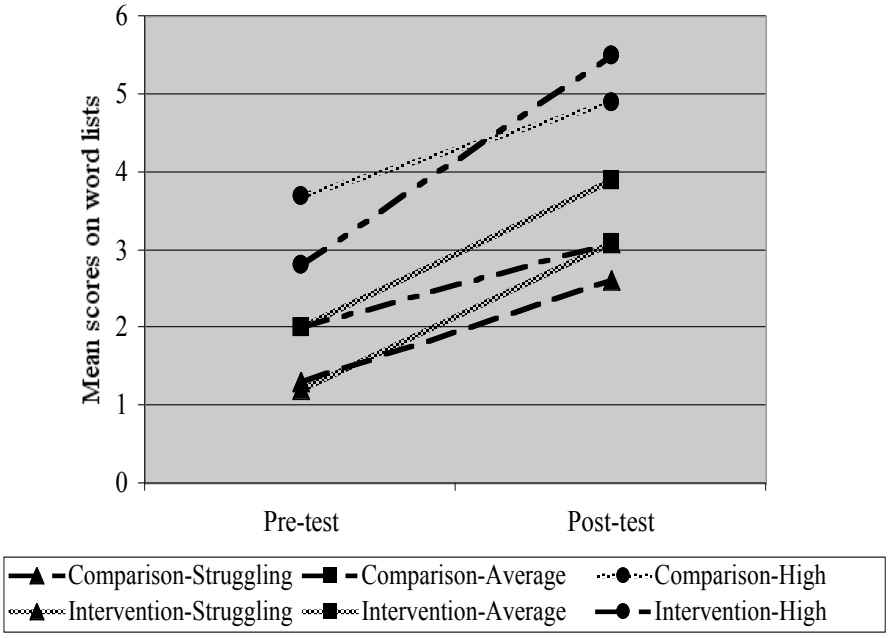

Figure 1. Increase in mean scores of children at different achievement levels on word-lists (pre- and post-test comparisons)

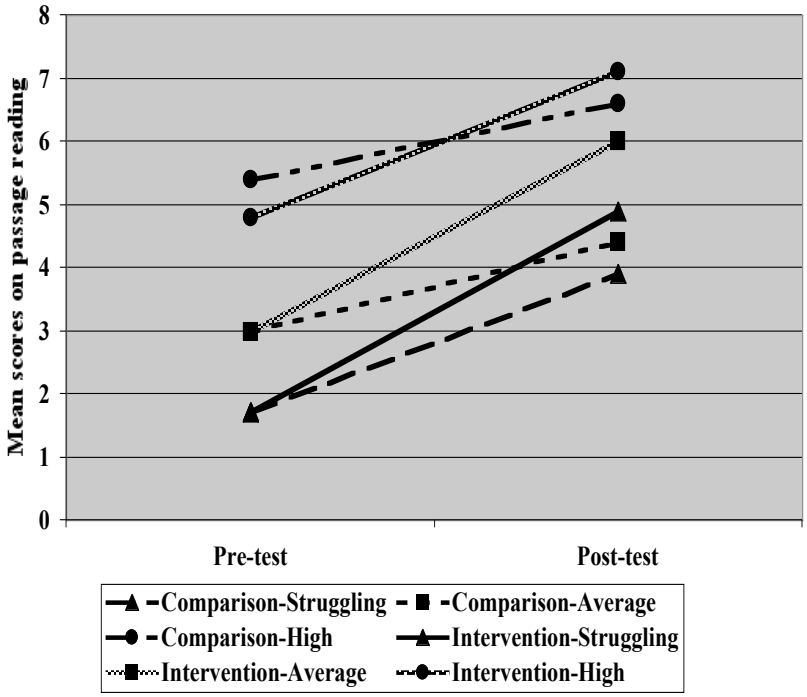

Figure 2. Increase in mean scores of children at different achievement levels on passage reading (pre- and post-test comparisons)