Investigating the instructional supportiveness of leveled texts

JAMES W. CUNNINGHAM
Emeritus, University of North Carolina at Chapel Hill, USA

STEPHANIE A. SPADORCIA
Lesley University, Cambridge, Massachusetts, USA

KAREN A. ERICKSON
University of North Carolina at Chapel Hill, USA

DAVID A. KOPPENHAVER
Appalachian State University, Boone, North Carolina, USA

JANET M. STURM
Central Michigan University, Mount Pleasant, USA

DAVID E. YODER
Emeritus, University of North Carolina at Chapel Hill, USA

For almost a century and a half, graded or leveled texts were generally considered an important, if not essential, component of elementary reading instruction (Hoffman, Roser, Salas, Patterson, & Pennington, 2001). Samuel Wood wrote the first graded readers early in the 1800s (Barr, 1989), and, by the turn of the 20th century, basal reading programs were often titled or described as “progressive..., not as a description of a ’new approach’ to teaching reading but as a description of the leveled nature of the books in the program” (Hoffman, Sailors, & Patterson, 2002, p. 270).

While some may have questioned the need for texts of gradually increasing difficulty, controversies centered on how, not whether, texts were to be sequenced or
LEVELED BOOKS originally selected by or produced for use in Reading Recovery or its regular classroom initiative are now also widely used in regular and special classrooms having no affiliation with Reading Recovery. The frequent use of these leveled books in settings other than Reading Recovery raises an important question: Do books leveled for use in Reading Recovery support other reading instructional emphases in addition to the ones that Reading Recovery teachers are trained to provide? The purpose of this study was to examine the curricular dimensions of books leveled for use in Reading Recovery in order to judge how supportive such texts are for early reading instruction emphasizing word recognition or decoding instead of, or in addition to, the three main cueing systems. The study found that Reading Recovery books, as a category of early reading instructional texts, provide only a moderate amount of support for word recognition instruction and almost none for decoding instruction in the use of onsets and rimes. The study also found that books leveled for use in Reading Recovery do not consistently increase in word-level demands as their levels increase.

LOS LIBROS nivelados, originalmente seleccionados o producidos para su uso en el programa Reading Recovery (Recuperación en Lectura) o en aulas regulares que adoptaban el programa, se usan actualmente en aulas regulares y especiales sin relación alguna con dicho programa. El uso frecuente de estos libros en contextos diferentes del de Reading Recovery introduce una pregunta importante: ¿los libros nivelados para Reading Recovery son adecuados en enfoques didácticos diferentes de aquellos para los que fueron capacitados los docentes de Reading Recovery? El propósito de este estudio es examinar las dimensiones curriculares de los libros nivelados de Reading Recovery con el fin de evaluar el apoyo que proporcionan esos textos en una didáctica de la lectura inicial que enfatiza el reconocimiento de palabras o la decodificación en lugar de, o además de, los tres principales sistemas de pistas. El estudio halló que, como libros de enseñanza de la lectura inicial, los textos de Reading Recovery proporcionan sólo un apoyo moderado a la enseñanza del reconocimiento de palabras y casi ninguno para enseñar decodificación con el uso de ataques y rimas. Asimismo el estudio encontró que en los libros de Reading Recovery no se halla un aumento de las demandas en el nivel de las palabras consistente con el aumento de los niveles.


ABSTRACTS

Investigating the instructional supportiveness of leveled texts

Investigando el apoyo didáctico brindado por textos nivelados

Ergründen anleitender Unterstützung durch eingestufte Texte
DES LIVRES standardisés par niveau, choisis ou produits à l’origine pour être utilisés dans le programme Reading Recovery (programme de rééducation de la lecture) ou à son initiative dans des classes ordinaires, sont maintenant utilisés également à grande échelle dans des classes ordinaires ou d’enseignement spécialisé non affiliées à Reading Recovery. L’utilisation fréquente de ces livres standardisés par niveau dans d’autres contextes que Reading Recovery soulève une question importante : des livres standardisés pour être utilisés en Reading Recovery peuvent-ils avoir une valeur instructive centre que celle que les enseignants de Reading Recovery sont formés à apporter ? Cette étude a pour but d’examiner l’intérêt par rapport aux programmes des livres standardisés par niveau pour être utilisés par Reading Recovery, afin d’évaluer l’intérêt de ces textes dans le cadre d’un enseignement des débuts de la lecture qui met l’accent sur la reconnaissance des mots ou le décodage au lieu de ou en plus des trois principaux systèmes d’indices. La recherche a montré que, en tant que textes pour l’enseignement des débuts de la lecture, les livres de Reading Recovery fournissent un apport modeste à l’enseignement de la reconnaissance des mots et presque aucun à l’enseignement du décodage en ce qui concerne l’utilisation des attaques et des rimes. La recherche a trouvé également que les livres standardisés par niveau pour une utilisation dans Reading Recovery n’augmentent pas les exigences au niveau mot au fur et à mesure que le niveau des textes s’élève.

ABSTRACTS

Examen de la valeur instructive de textes standardisés par niveau
leveled. In the mid-1980s, however, a sea change occurred in mainstream basal readers marked by the lack of “any systematic attention to the decoding demands of the texts” (Hoffman et al., 2002, p. 272). Soon, reading educators from different perspectives were expressing concern over the near abandonment of appropriate text difficulty as a principle in elementary reading instruction. Clay (1991) observed,

> There is an exciting enthusiasm among teachers in some countries today for teaching from story books but this is often associated with a strong disregard for any gradient of difficulty in the texts used. Any levelling of books is seen to be unnecessary and an impediment to learning.... However, many children learning to read will be confused without assistance from some form of a gradient of difficulty in reading books. (p. 201)

In a similar manner, Chall, Conard, and Harris-Sharples (1991) summarized a number of recent articles and books:

> The common theme of these professional articles was that the difficulty level of [children’s] textbooks was an unimportant issue. (p. 27)

> It has...become popular to say that [children’s] textbooks are suitable if they are interesting and contain high-quality writing, and that it matters little how difficult they are. (p. 107)

During the time that the literature-based and whole language movements were bringing the principle of gradually increasing text difficulty into question, the Reading Recovery program was simultaneously gaining a foothold in the United States. The principle of gradually increasing text difficulty in small increments has been central to the Reading Recovery early intervention program from the beginning (Clay, 1985, 1991; Peterson, 1991; Pinnell, 1990). Its application of the principle can be seen as consistent with the long-standing practice of determining a child's reading instructional level (Betts, 1946), although Reading Recovery uses different procedures and criteria when leveling books and matching them to readers (Peterson).

Curricular dimensions of increasing text difficulty

Although the principle of gradually increasing text difficulty has historically, and again recently, been widely considered crucial for early reading instruction, this consensus does not extend to the characteristics the sequence of texts should have. Rather, most advocates of different approaches to teaching early reading have argued that texts should gradually increase in difficulty along the dimension(s) of emphasis in the instructional approach they promote. Three major dimensions of increasing challenge have been emphasized in texts selected or written to complement various approaches to early reading instruction.

Text with vocabulary control by word frequency

In the 1920s and 1930s, the “vocabulary control movement” began in foreign language learning. While the first words to be learned in a second language can be selected on any basis, the vocabulary control movement was soon dominated by those arguing that more successful language learning is likely to occur if the vocabulary of instruction generally consists of the highest frequency words (Schmitt, 2000). In reading education,

> One prominent characteristic of the readers of [the 1925–1935] period was the use of standard word lists as a basis for selecting the vocabulary. Writers of textbooks took meticulous care to have the vocabulary in their primers and first readers consist almost wholly of words having the highest frequency. (Smith, 1965/2002, p. 203)

By the 1950s, most basal reading series controlled the frequency, number, and repetition of words in their selections, particularly at the lower levels (Hoffman et al., 2001; Smith, 1965/2002). This vocabulary control by word frequency was the main tool for sequencing the texts and assigning levels to the books in these widely used series (Hoffman et al., 2002). Vocabulary control based primarily on word frequency continued to hold sway in the design of primary materials in mainline basal series through the mid-1980s (Hiebert, 1999).

In texts with traditional vocabulary control, as text difficulty increases, the percentage and repetition of high-frequency words decreases or the list of words considered familiar due to frequency range increases in size. Either way, vocabulary control is loosened as texts become more difficult. The purpose of vocabulary
control by word frequency is to foster the students’ word recognition—the ability to access pronunciations attached in memory to entire printed words. The instructional emphasis the texts are designed or selected to support is word study instruction—guiding students’ work with specific high-frequency words to help them learn the particular pattern of letters that comprise each one (Cunningham, Koppenhaver, Erickson, & Spadorcia, 2004).

**Text with phonetic control**

**Decodable text**

Phonetic control increases the percentage of words in a text selected for their word decodability (Juel & Minden-Cupp, 2000). Word decodability always includes attention to phonetic regularity—the predictability of the words’ letter-sound patterns—and the number of syllables (Juel & Roper/Schneider, 1985; Martin & Hiebert, 1999; Menon & Hiebert, 1999). In addition, word decodability may also include consideration of a perspective attributable to Beck (1981): “Within this [instructional consistency] perspective the decodability of a word is determined by the instruction that has preceded the appearance of the word in a selection” (Hoffman et al., 2002, p. 276).

As the relative difficulty of texts with phonetic control increases, the percentage of decodable words decreases or the complexity of words considered decodable increases. Either way, phonetic control is loosened as texts become more difficult (Juel & Roper/Schneider, 1985). The purpose of phonetic control is to foster students’ decoding—their ability to use knowledge of letter-sound relationships and patterns to construct probable pronunciations for unfamiliar printed words. The instructional emphasis the texts are designed or selected to support is coaching instruction—guiding students during oral reading to use strategies and multiple sources of information to solve problems as they arise (Hicks & Villaume, 2000).

**Predictable text**

We use the term **predictable text** to include all written material selected or constructed to provide extra cues to readers so they can read it accurately, even though their word recognition and decoding abilities are not adequate to identify a satisfactory percentage of the words presented in isolation. That is, predictable text is printed language chosen or engineered to amplify the availability of context clues of various kinds. Some sources of predictability include the match between illustrations and print, the familiarity of language patterns and story episodes (Peterson, 1991), as well as rhyme and repeated phrases (Hoffman et al., 2002).

As the relative difficulty of a sequence of predictable texts increases, the availability of extra context clues decreases. Texts become “longer,” “more complex,” and have “less patterned language and more varied vocabulary” (Peterson, 1991, p. 123). That is, predictability declines as texts become more difficult. The purpose of predictable text is to foster students’ coordinated use of the three main cueing systems—“sentence structure (syntax),” “message (semantics),” and “letters (graphic cues)” (Clay, 1993, pp. 41–42). The instructional emphasis the texts are designed or selected to support is coaching instruction—guiding students during oral reading to use strategies and multiple sources of information to solve problems as they arise (Hicks & Villaume, 2000).

**Multiple-criteria text**

Hiebert (1999) referred to text with a dominant principle such as decodability or vocabulary control by word frequency as a “single-criterion text” (p. 563). She then described how teachers can use texts with different single criteria to provide a “multiple-criteria program” (p. 563). She also expressed the hope that publishers will follow the example of Dr. Seuss and create “multiple-criteria texts” (p. 563). Presumably, if such texts were leveled, multiple-criteria texts would manifest declines in vocabulary control by word frequency, phonetic control, and predictability as their assigned levels increased.

**Rationale for the study**

Little is known about the contribution of texts to early reading instruction. One reason may be an attitude toward texts of effectiveness by association. For example, Beck (1997) argued that, because highly decodable text is so closely associated with synthetic phonics instruction, research supporting the latter supports the former. If such an attitude were common among researchers of early reading, it would help explain why so much more research exists on instructional approaches than on instructional texts.

Another reason that so little is known about the contribution of texts to early reading instruction may be the lack of an accepted technology for validating the schemes used to assign relative or absolute difficulty to the texts. Readability research has proven unable to account for the small gradations of difficulty that
mark the typical sequencing or leveling of early reading materials (Hoffman et al., 2001; Klare, 1984). As a consequence, research on texts for early reading instruction that depends on the validity of the levels of those texts is less likely to be done or published.

This study examined the instructional supportiveness of texts for teaching early reading. We define instructional supportiveness as the degree to which texts provide opportunities for children to apply what they are being taught and gradually increase the challenge of those applications. Research on instructional supportiveness proceeds by analyzing curricular dimensions of one or more sets of sequenced texts.

Curricular dimensions of any set of sequenced texts are manifest in two ways that have not been previously distinguished in the literature. The first one is the opportunity for readers to apply one or more instructional emphases across the texts as a set (curricular dimensions of the set of texts as a category of reading instructional texts). The second one is the factors that become more demanding as the books in the set increase in difficulty (curricular dimensions of the set’s leveling). These two ways correspond to the two aspects of instructional supportiveness: the opportunities for students to apply what is taught, and the gradual increase in challenge of those applications.

Investigating the instructional supportiveness of texts avoids the assumption of effectiveness (or ineffectiveness) by association. That is, it brackets the question of whether a particular reading instructional approach is effective or best. Instead, it empirically examines the match between curricular dimensions of sets of texts and one or more instructional emphases. To the extent that “text matters in learning to read” (Hiebert, 1999, p. 552), any reading instructional approach can be well or poorly served by the degree to which the texts that accompany it provide increasingly challenging opportunities for the children to apply what they are being taught in that approach. As a consequence, investigating the instructional supportiveness of texts has the potential to improve the effectiveness of different reading instructional approaches without waiting for “the reading wars” to be settled.

Although the lack of a technology for validating text levels may have hindered research on instructional texts in the past, that obstacle can be avoided by examining instructional supportiveness. A set of leveled texts is analyzed for evidence that it supports one or more particular instructional emphases. The assigned levels of the texts examined are not assumed to be valid. If evidence is found that the leveling of the texts supports a particular instructional emphasis, that finding provides some validation for the texts’ levels, but only in support of that particular instructional emphasis. The texts still may not be validly leveled for any other instructional approach or with regard to their difficulty for readers in general on, say, a cloze task or an oral reading fluency task.

Determining the instructional supportiveness of texts is prerequisite to determining their instructional value. Ascertaining the contribution of a type of text to reading achievement requires intervention research that varies texts (contrasting different kinds of texts, or texts versus no texts) without varying the instruction itself (Allington & Woodside-Jiron, 1998). However, without an analysis of the texts in the intervention for their supportiveness of the instruction provided, the results of the intervention study are uninterpretable. For example, a significant effect for texts could occur because they support an instructional emphasis other than that in the intervention, but that some participants have or had outside the intervention. Or the effect for texts may not be significant because they fail to provide enough opportunities for participants to apply what they were being taught in the intervention with gradually increasing challenge.

In this study, we investigated the instructional supportiveness of Reading Recovery books for two reasons. First, research on the instructional supportiveness of texts, as a new research approach, overcomes the current obstacles to researching the contribution of texts to early reading instruction. The role that instructional texts play is an important area of reading education where knowledge is both lacking and needed. Second, this specific study is needed because books originally leveled for use in Reading Recovery are now also widely used in regular and special classrooms having no affiliation with Reading Recovery. Apparently, these “little books” have filled the gap left when many publishers of reading instructional materials downplayed the need for a continuum of difficulty between the mid-1980s and the mid-1990s.

Currently, many teachers and publishers seem to have become committed not only to the principle of gradually increasing text difficulty but also to the accuracy of Reading Recovery levels themselves for all reading instructional settings. In many elementary schools, the Reading Recovery level of a book appears to determine the perceived difficulty of that book for all purposes, whether instruction or assessment, and regardless of instructional setting—whether regular classroom, Reading Recovery, special education, Title I reading, or volunteer tutoring (Title I is a U.S. federally funded program for at-risk students). However, the frequent use of these leveled books in settings
other than Reading Recovery and its affiliated programs raises an important question for research: Do books leveled for use in Reading Recovery support other reading instructional emphases besides the use of the three main cueing systems?

Method

The purpose of this study was to examine books leveled for use in Reading Recovery in order to judge how supportive such texts are for early reading instruction emphasizing word recognition or decoding instead of, or in addition to, the emphasis in Reading Recovery on the three main cueing systems.

Reading Recovery levels

Books are leveled by the Reading Recovery Council of North America on the basis of how well they “support the reader’s present knowledge and, at the same time, provide some new challenge and opportunity for engaging in ‘reading work’” (Peterson, 1991, p. 123). “Present knowledge” includes what the reader has learned from previous reading instruction, both in and out of Reading Recovery, as well as the knowledge of language, literacy, and life that the child brings to reading instruction (Peterson).

Reading Recovery leveling was the first of several schemes that share many of the same conceptual bases and lead to levels that more or less correspond with one another. For example, Fountas and Pinnell (1996) explained that their leveling is very similar to Reading Recovery’s, but that their levels are less fine-grained and receive ascending letters rather than numbers. Three of the most popular schemes that yield book levels similar to Reading Recovery are Fountas and Pinnell leveling, Wright Group/McGraw Hill leveling, and Developmental Reading Assessment leveling (Wright Group, 2004). We chose to use Reading Recovery leveling in this study because of its longevity, wide use, and the large network of users of the book list who have input on the leveling and releveling of books. We believe the results of our study would be comparable had we conducted it with books leveled by any of the schemes conceptually related to Reading Recovery leveling.

Materials

To select the texts for this study, we used the master list of books leveled by Reading Recovery (Reading Recovery Council of North America, 1997). The Reading Recovery early intervention program levels trade books and sets of books produced for beginning readers by educational publishers into 20 levels (Peterson, 1991; Reading Recovery Council of North America). Level 1 books are those that kindergartners can learn to “read” during their first semester; level 20 books are those considered readable by average second graders early in the school year.

We systematically selected four books at each of the 20 levels for analysis. The books were not randomly selected because we had difficulty locating so many of the books on the list. Basically, our strategy was to use public libraries, large bookstores, and Reading Recovery teachers in our local area until we had obtained four of the books at each level. We did not exclude any book on the master list. This process resulted in a sample of 80 books.

Measures

To our knowledge, this was the first study ever done on the instructional supportiveness of text. As a consequence, there were no established measures of the construct for us to use. Before we could proceed with our investigation of the instructional supportiveness of leveled texts, we had to determine measurement principles for the construct. We followed four measurement principles in developing and selecting the specific measures for this study.

1. Different factors contribute to the instructional supportiveness of text. In this study, we were interested in print-based rather than book-design or reader-based factors. For us, the challenge a text will present to a particular student at a specific time is dependent upon print-based, book-design, reader-based, and situational factors. Investigations of the instructional supportiveness of text can focus on one or more of these factors. In this study, we restricted our attention to print-based factors because we wanted to investigate the supportiveness of leveled texts for print-based reading instruction.

2. Measures of the instructional supportiveness of text should respect the multilevel nature of texts. The construct of instructional supportiveness of text cannot be measured by examining only one aspect of text. Otherwise the construct would be the instructional supportiveness of items, whether they be words, sentences, texts, or something else. While controversies remain, most “theorists distinguish among component processes of reading at the word, sentence, and text level” (Haberlandt, 1994, p. 2). In keeping with this consensus, any study of instructional supportiveness of text should include at least three kinds of measures corresponding to these three levels of text structure: word, sentence, and discourse (text). Consistent with this principle, our main consideration in this study was to choose measures that focus on each major level of text structure, not just the word level.
3. Multiple measures of each aspect of instructional supportiveness are best. Because there are no established measures of the construct of the instructional supportiveness of leveled texts, we concluded that it was important to have several measures at each of the three major levels of text structure, rather than trying to choose, in advance of data, the single best measure at each one. Any selection of a single best measure for a level of text structure should be based on the results of empirical analyses and will be limited to the set of texts analyzed or the category of texts that set represents. Because of this principle, we chose several measures at each level of text structure for this study.

4. Each measure should reflect current best assessment practices in quantifying the demands at each level of text structure. Consistent with this principle, each of the measures in this study was chosen or developed to represent the best measurement technology currently available for capturing the print-based demands at each of the three major levels of text structure.

We contend that these four measurement principles should guide the selection of measures for any study of the instructional supportiveness of text, regardless of the specific research question. The research question only influences which word-level, sentence-level, and discourse-level demands will be measured using multiple measures that are among the best currently available.

Measures of word-level demands. This study was designed to answer the research question: “Do books leveled for use in Reading Recovery support other reading instructional emphases besides the use of the three main cueing systems?” Several approaches were taken to obtain measures of the word-level demands of texts that support instruction emphasizing word recognition or decoding instead of, or in addition to, an emphasis on the three main cueing systems. First, five measures were chosen to examine different aspects of vocabulary control by word frequency. These measures, except type-token ratio, were based on statistics from The Educator’s Word Frequency Guide (Zeno, Ivens, Millard, & Duvvuri, 1995), a large recent corpus of the words in printed English, and were computed technologically using their CD-ROM. Type-token ratio was also generated by computer.

*Mean U of the words.*—The U statistic is the frequency of a word per million running words of text, weighted by a measure of how widely distributed that frequency is across texts from different subject areas (Carroll, 1971). Larger U’s indicate higher frequency with wider distribution. The mean U of the unique words (types) in a book is a measure of word frequency that takes every different word into account.

*Percentage of high-frequency words (100-word list).*—The percentage of the running words in a book that are the 100 most common words in printed English.

*Percentage of high-frequency words (500-word list).*—The percentage of the running words in a book consisting of the 500 most common words and their regular morphological variations.

*Type-token ratio.*—The number of unique words (types) in a book divided by the number of running words in the book (tokens). This is a measure of how much word repetition occurs in a book.

*U of the word ending the most frequent 75% of the words.*—This measure was determined by first sorting the unique words (types) in the book by their U statistics. Then we drew a line separating the 75% of the words with the highest U’s from the 25% of the words with the lowest U’s. We then selected the word in the larger set that had the lowest U (i.e., was closest to the line). This is a measure of the minimum frequency of the 75% of the most frequent unique words in the book. The U of the least frequent word of the most frequent 75% of the words in a book is a measure of word frequency that is less sensitive to the presence of a few words with extremely low U’s such as some proper names.

The four measures employing the U statistic are based on the lognormal model of word-frequency distribution (Carroll, 1971) as applied in a recent corpus of words in printed English (Zeno et al., 1995). The fifth of these measures, type-token ratio, was recently used by Hiebert (1999) to indicate the amount of repetition a text provides on the words it contains.

Second, we developed three measures of onsets and rimes decodability. We had no reason to expect that Reading Recovery books are leveled to support synthetic phonics instruction. However, because predictable text often includes rhyme (Hoffman et al., 2002), we did have reason to expect that books leveled for use in Reading Recovery may support instruction in the use of onsets and rimes to decode one-syllable, phonetically regular words.

*Percentage of onsets and rimes decodable words (list A).*—The percentage of the running words in a book on a list of the one-syllable words comprised of the highest utility onsets and rimes (onsets and rimes with the largest neighborhood sizes in previous studies). In addition, for this study, rimes that were themselves among the 100 highest frequency words were also combined with the highest utility onsets. List A consisted of such words as bug, hat, and lap.

*Percentage of onsets and rimes decodable words (list B).*—The percentage of the running words on list A plus one-syllable words comprised of high- to moderate-utility onsets and rimes. List B added such words to list A as dill, clock, and broke.

*Percentage of onsets and rimes decodable words (list C).*—This measure was the percentage of the running words on list B plus one-syllable words comprised of moderate-utility onsets and rimes. List C added such words to list B as trail, yam, and trend.
Most studies of onset-rime decoding have been based on neighborhood counts (number of words containing a particular onset or rime) to determine utility (e.g., Goswami, 1998; Leslie & Calhoon, 1995). The rime utilities used to select the rimes in these three measures were from the two major studies of orthographic rime neighborhood sizes (Fry, Fountoukidis, & Polk, 1985; Wylie & Durrell, 1970). The onset utilities used in this study were from a recent study conducted as part of a study on first and second graders’ onset-rime decoding (Cunningham et al., 1999). The items in the three tests also came from that study, and the division of the items into three subtests was based on the results of that study.

Third, because word identification should lead to word meaning access, we selected one more measure of word-level demands:

*Morphem per 100 words*—The morpheme is the smallest linguistic unit of meaning. The word, *falling*, for example, consists of the free morpheme, *fall*, and the bound morpheme, *ing*. This measure was computed by counting the number of morphemes in each book, dividing by the number of words, and multiplying the result by 100.

This last measure of word-level demands was also seen as a decodability measure, albeit of a grapho-semantic rather than grapho-phonetic kind. It was included in this study because of renewed interest in the morpheme as another multiletter unit of decoding besides onsets and rimes (Ehri, 1998; Minkoff & Raney, 2000).

These nine measures were chosen to indicate whether leveled texts have word-level demands that support instruction that teaches students to recognize high-frequency words or that teaches them to decode unfamiliar words comprised of high-utility onsets and rimes.

*Measure of sentence-level demands.* Sentence-level factors, principally syntax, constitute a major component of the context within which word identification occurs (Graesser & Bertus, 1998; Graesser, Swamer, Baggett, & Sell, 1996). While there are competing theories for the relationship between word identification and context during proficient reading, we could find no advocate of an instructional emphasis on high-frequency words or phonics in the literature who also advocated using only isolated words rather than texts for students to apply what they were learning. Indeed, effective teachers who emphasize word recognition and decoding the most are equally concerned with teaching their students to apply these words and skills to reading connected text (Taylor, Pearson, Clark, & Walpole, 2000).

In this study, two approaches were taken to generate measures of the sentence-level demands of leveled texts to indicate whether such texts are supportive of instruction that emphasizes high-frequency words or phonics. First, we chose two measures of sentence complexity:

*Morphem per sentence*—The number of morphemes in the book divided by the number of sentences.

*Words per sentence*—The number of running words in the book divided by the number of sentences.

According to Miller and Kintsch (1980), average sentence length in a text predicts the number of processing cycles required of a reader by the sentences in a text. In addition, because we had no reason to believe that either measure was in the minds of the authors of books leveled for use in Reading Recovery, they could be expected to predict the general sentence complexity of those books.

Second, we selected two parallel measures of syntactic complexity that were based on the T-unit rather than the sentence. Hunt (1964) proposed dividing subjects’ writing samples into repunctuated sentences he defined as “minimal terminable units...the shortest grammatically terminable units into which a connected discourse can be segmented without leaving any fragments as residue... For short, the term T-unit will be used” (p. 34). Each T-unit is an independent clause with all subordinate clauses attached to it.

*Morphem per T-unit*—The number of morphemes in the book divided by the number of T-units.

*Words per T-unit*—The number of running words in the book divided by the number of T-units.

T-unit length is currently used as an indicator of both receptive and expressive language syntactic maturity/complexity in research on second-language acquisition (Gass & Selinker, 2001) and in speech and hearing sciences (Paul, 2001).

These four measures were chosen to indicate whether leveled texts have sentence-level demands supporting instruction that emphasize word recognition or decoding.

*Measure of discourse-level demands.* Finally, five measures of discourse-level demands were chosen, based on previously discussed linguistic units:

*Number of morphemes in the book*
*Number of running words (tokens) in the book*
*Number of sentences in the book*
*Number of T-units in the book*
*Number of unique words (types) in the book*
In every case, the measure of discourse-level demands was an estimate of overall text difficulty in terms of the total amount of word- or sentence-level processing required to read the book. Two of the measures, number of sentences in the book and number of T-units in the book, were estimates of the total amount of sentence-level processing required by the book. The other three measures were estimates of the total amount of word-level processing required by the book.

Each of these measures was chosen for one of two reasons. The total amount of word-level processing in a book should indicate the degree of automaticity or fluency of word identification (Samuels, 2002) required to sustain success throughout the book. The total amount of sentence-level processing should indicate the relative ease or difficulty of integrating word identification with context throughout the book (Kintsch, 1988/1994).

These 18 measures were chosen as the best ones available for indicating whether leveled texts have word-, sentence-, and discourse-level demands that support instruction that teaches students to recognize high-frequency words or that teaches them to decode unfamiliar words comprised of high-utility onsets and rimes.

**Procedure**

Scores on all 18 measures of print-based demands were computed for each of the 80 books in the sample. All the print in a book, exclusive of front matter, end matter, and illustrations, was analyzed to yield each score for that book. We planned to examine the means and standard deviations of these measures across the 80 books to determine characteristics of Reading Recovery books as a category of instructional texts for early readers. We also planned a series of simple and multiple regression analyses using the 18 measures to predict the levels of the 80 books to explore Reading Recovery leveling.

**Results**

Means and standard deviations were computed for each of the 18 measures of print-based demands. See Table 1 for these means and standard deviations. Each of the 18 measures was also correlated with the Reading Recovery level of the 80 books. We tested each of the correlations for whether it indicated that the measure was a significant predictor of increasing curricular demands at higher Reading Recovery levels. This testing required a two-step process. The first step was to determine the sign (+ or -) of each correlation that would indicate an increased challenge on that measure as Reading Recovery level increased. The second step was to test the statistical significance of any correlation having a sign that indicated increasing challenge with increasing level.

Before computing the correlations, we assigned to each of the 18 measures the sign of its correlation

<table>
<thead>
<tr>
<th>Measure</th>
<th>$M$</th>
<th>$SD$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Word-level demands</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean $U$ of the words</td>
<td>7918.4</td>
<td>4383.7</td>
</tr>
<tr>
<td>Morphemes per 100 words</td>
<td>109.9</td>
<td>19.8</td>
</tr>
<tr>
<td>Percentage of high-frequency words (100-word list)</td>
<td>46.1</td>
<td>18.4</td>
</tr>
<tr>
<td>Percentage of high-frequency words (500-word list)</td>
<td>58.9</td>
<td>19.1</td>
</tr>
<tr>
<td>Percentage of onset-rime decodable words (list A)</td>
<td>9.3</td>
<td>7.2</td>
</tr>
<tr>
<td>Percentage of onset-rime decodable words (list B)</td>
<td>12.1</td>
<td>9.0</td>
</tr>
<tr>
<td>Percentage of onset-rime decodable words (list C)</td>
<td>16.0</td>
<td>11.4</td>
</tr>
<tr>
<td>Type-token ratio</td>
<td>0.4</td>
<td>0.2</td>
</tr>
<tr>
<td>$U$ of the word ending the most frequent 75% of the words</td>
<td>171.4</td>
<td>238.6</td>
</tr>
<tr>
<td><strong>Sentence-level demands</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Morphemes per sentence</td>
<td>10.5</td>
<td>9.0</td>
</tr>
<tr>
<td>Morphemes per T-unit</td>
<td>7.0</td>
<td>3.7</td>
</tr>
<tr>
<td>Words per sentence</td>
<td>9.6</td>
<td>7.6</td>
</tr>
<tr>
<td>Words per T-unit</td>
<td>6.4</td>
<td>3.3</td>
</tr>
<tr>
<td><strong>Discourse-level demands</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of morphemes</td>
<td>227.5</td>
<td>331.3</td>
</tr>
<tr>
<td>Number of running words (tokens)</td>
<td>214.2</td>
<td>313.4</td>
</tr>
<tr>
<td>Number of sentences</td>
<td>28.4</td>
<td>45.1</td>
</tr>
<tr>
<td>Number of T-units</td>
<td>33.5</td>
<td>42.8</td>
</tr>
<tr>
<td>Number of unique words (types)</td>
<td>70.1</td>
<td>72.8</td>
</tr>
</tbody>
</table>

**TABLE 1**

MEANS AND STANDARD DEVIATIONS OF 18 MEASURES OF PRINT-BASED DEMANDS
with Reading Recovery level that would indicate increasing print-based demands as book level increased. Ten of the 18 measures were tagged with a plus as the sign of the correlation that would indicate increased challenge on the measure at higher levels. The remaining eight measures were tagged with a minus as the sign indicative of such an association. For example, type-token ratio and words per T-unit were two of our 18 measures with different signs tagged to them. Type-token ratio was expected to correlate negatively with book level if Reading Recovery leveling provides students with more repetition in identifying the same words at lower levels and less repetition at higher levels. This was so because higher type-token ratios indicate more repetition of the same words. However, words per T-unit was expected to correlate positively with book level if Reading Recovery leveling requires students to read less sophisticated syntax at lower levels and more sophisticated syntax at higher levels. This was so because longer T-units indicate more sophisticated syntax. (See Table 2 for the signs tagged to each measure.)

After the first step of the testing procedure was performed, 10 of the 18 correlations were found to have the sign indicating increasing print-based demands as Reading Recovery level increases. In the second step, the 10 correlations with the tagged sign were tested for statistical significance. Because the first step had determined the signs of the correlations to be tested for statistical significance, these significance tests were one-tailed. Seven of the 10 correlations were statistically significant ($p < .05$).

However, because 10 significance tests had been performed, there was an increased danger of making a Type I error. So, to avoid the multiple-testing fallacy, Bonferroni’s correction of the 10 alpha levels was performed and the 10 correlations examined again for significance. Bonferroni’s correction requires that the alpha level for 10 tests be set at $p < .005$ in order to achieve an actual alpha level of $p < .05$. The same 7 correlations remained significant after Bonferroni’s correction was performed. (See Table 2 for the 18 simple correlations, tagged signs, and statistical significance of correlations with the tagged sign, both before and after Bonferroni’s correction.)

Our plan was to produce two hierarchical multiple regression equations to predict the books’ Reading Recovery levels in order to (a) evaluate how much of the increase in Reading Recovery level is determined by increasing print-based demands and to (b) deter-

### TABLE 2
**CORRELATIONS WITH READING RECOVERY LEVEL OF 18 MEASURES OF PRINT-BASED DEMANDS**

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Sign indicating increasing challenge at higher levels</th>
<th>$r$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Word-level demands</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean $U$ of the words</td>
<td>-</td>
<td>.21</td>
</tr>
<tr>
<td>Morphemes per 100 words</td>
<td>+</td>
<td>.06</td>
</tr>
<tr>
<td>Percentage of high-frequency words (100 word list)</td>
<td>-</td>
<td>.35</td>
</tr>
<tr>
<td>Percentage of high-frequency words (500 word list)</td>
<td>-</td>
<td>.12</td>
</tr>
<tr>
<td>Percentage of onset-rime decodable words (list A)</td>
<td>-</td>
<td>.02</td>
</tr>
<tr>
<td>Percentage of onset-rime decodable words (list B)</td>
<td>-</td>
<td>.09</td>
</tr>
<tr>
<td>Percentage of onset-rime decodable words (list C)</td>
<td>-</td>
<td>.04</td>
</tr>
<tr>
<td>Type-token ratio</td>
<td>-</td>
<td>.04</td>
</tr>
<tr>
<td>$U$ of the word ending the most frequent 75% of the words</td>
<td>-</td>
<td>-.13</td>
</tr>
<tr>
<td><strong>Sentence-level demands</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Morphemes per sentence</td>
<td>+</td>
<td>-.03 ♠</td>
</tr>
<tr>
<td>Morphemes per T-unit</td>
<td>+</td>
<td>.53*†</td>
</tr>
<tr>
<td>Words per sentence</td>
<td>+</td>
<td>.01</td>
</tr>
<tr>
<td>Words per T-unit</td>
<td>+</td>
<td>.54*†</td>
</tr>
<tr>
<td><strong>Discourse-level Demands</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of morphemes</td>
<td>+</td>
<td>.46*†</td>
</tr>
<tr>
<td>Number of running words (tokens)</td>
<td>+</td>
<td>.47*†</td>
</tr>
<tr>
<td>Number of sentences</td>
<td>+</td>
<td>.40*†</td>
</tr>
<tr>
<td>Number of T-units</td>
<td>+</td>
<td>.36*†</td>
</tr>
<tr>
<td>Number of unique words (types)</td>
<td>+</td>
<td>.69*†</td>
</tr>
</tbody>
</table>

*Correlation with both the expected sign (+ or -) and $p < .05$, one-tailed test.
†Correlation with both the expected sign (+ or -) and $p < .05$, one-tailed test, after Bonferroni’s correction.
mine which of the print-based demands—word-, sentence-, or discourse-level—was most associated with the gradient of difficulty considered important for texts to have that supports their reading instruction.

To increase the likelihood of cross-validation of these equations, we had planned for the first equation to have only three predictors—the best significant predictor of increasing demands of each of the three kinds, in the order of size of simple correlation. We had planned for the second equation to have only four predictors—the best significant predictor of increasing demands of each of the three kinds plus the best significant predictor of increasing word-level demands of the other kind (decodability or word frequency) than was included in the first equation. In other words, we intended to include both decodability and word-frequency measures in the second equation after letting either one of the two serve as the best measure of increasing word-level demands in the first equation. Again, in the second equation, the best predictors of each of the four types would be entered in the order of size of simple correlation. To produce the second equation, percentage of onset-rime decodable words (lists A–C) and morphemes per 100 words were considered to be word decodability measures. The rest of the measures of word-level demands were considered to be word-frequency variables. Each equation was to be cross-validated by randomly partitioning the books into two sets of equal size.

Based on the simple correlations, the best single predictor overall of increasing curricular demands as Reading Recovery level increases was number of unique words (types) in a book. This predictor was placed into both planned multiple regression equations as the first predictor. Because this variable is a measure of increasing discourse-level demands, no other measure of increasing discourse-level demands was considered for either planned equation.

The best predictor of increasing challenge of the other two kinds, word-level demands or sentence-level demands, and the second best predictor of increasing demands overall, was words per T-unit. This variable was placed into both planned multiple regression equations as the second predictor. Because this variable is a measure of increasing sentence-level demands, no other measure of increasing sentence-level demands was considered for either planned equation.

Unfortunately, no measure of increasing word-level demands was significant. As a consequence, no measure of increasing word-level demands was added to either planned multiple regression equation. Our intention to produce two multiple regression equations was thwarted. Instead, we were left with a single equation having two predictors, the first a measure of increasing discourse-level demands and the second a measure of increasing sentence-level demands. Using this single set of two predictors, a multiple regression analysis with Reading Recovery level as the criterion was performed on the 80 books.

The multiple correlation between the set of two predictors and Reading Recovery level was .78 ($p < .001$), accounting for 60% of the variance in Reading Recovery level. The first predictor accounted for 47% of the variance, while the second accounted for 13% ($p < .01$). A multiple correlation of .78 lies in the range of .7 to .9 that has been traditional for most readability formulas predicting the difficulty of texts with known difficulty (Chall & Dale, 1995). The multiple regression equation was

$$\text{Predicted Reading Recovery level} = 3.011 + .0465 (\text{number of unique words in the book}) + .663 (\text{words per T-unit})$$

This equation was cross-validated. The four books at each of the 20 levels were randomly assigned to two subsets of 40 books. One of these two sets was randomly selected and a multiple regression analysis was performed on those 40 books using the same two predictors in the same order: number of unique words (types) and words per T-unit. This analysis produced a multiple regression equation based on the data from those 40 books ($R = .82$). The constant and beta weights from this analysis were then used to compute a predicted Reading Recovery level for each of the 40 books in the other subset. Finally, a simple correlation was computed between the predicted and actual Reading Recovery levels for the 40 books in the second subset. This correlation ($r = .73$) was compared with the multiple correlation of the first subset of books as the traditional test of cross-validation. The decline in the correlation was less than .1 (.08), supporting the probable stability of the multiple regression equation.

Discussion

Limitations of the study

The 80 books we systematically selected as our sample may not have been representative of all the Reading Recovery books available for use with children at any one time. Replications of this study with other samples of books leveled for use in Reading Recovery would be necessary to determine how representative our sample was.
One of our measurement principles was that each measure in this study should reflect the current best assessment practices in quantifying the demands at each level of text structure. Even so, there may be measures other than the 18 we used that would be better indicators of the print-based demands at the word, sentence, and discourse levels of books leveled for use in Reading Recovery, which indicate how supportive such texts are for early reading instruction that emphasizes word recognition or decoding instead of, or in addition to, the emphasis in Reading Recovery on the three main cueing systems. Or better measures may emerge in the literature in the future. The findings of this study are limited to the set of measures we used.

**Instructional supportiveness of the category of Reading Recovery books for instruction emphasizing word recognition or decoding**

When means and standard deviations of the 18 measures of print-based demands were interpretable as curricular dimensions, they were examined to characterize the instructional supportiveness for word recognition or decoding of the 80 books as a whole, without regard to how the books are leveled. To the extent that our sample represented Reading Recovery books in general, these characterizations can be said to apply to Reading Recovery books as a category of instructional texts for early reading.

**Word-level demands.** We conclude that Reading Recovery books in general contain a moderate proportion of high-frequency words. In the average Reading Recovery book in our sample, 46.1% of the running words were from the 100 most frequent words in printed English. These highest frequency words appear slightly less often in Reading Recovery books than they do in texts written at a third-grade level where half the running words (tokens) are made up of the 100 most frequent types (Adams, 1990; Carroll, Davies, & Richman, 1971). In a similar manner, in the average Reading Recovery book in our sample, 58.9% of the running words were from the 500 most frequent words in printed English. These high-frequency words appear less often than would be expected, given that “only 300 different words and their variants make up 65 percent of all written material [at third-grade level and beyond]” (Sakiey & Fry, 1979, p. 2).

The type-token ratio for words in the average book in our sample was .4 (1 to 2.5). Hiebert (1999) discussed the instructional role of type-token ratio or “word density ratio of unique to total words” (p. 560). She contended that the type-token ratios of 1:2 or 1:4 in the little books she examined probably do not provide children with adequate repetition on the words they contain. Given that standard, the Reading Recovery books in our sample as a whole do not provide an adequate amount of repetition on the high-frequency words.

Unfortunately, there are no guidelines for percentage of high-utility onset-rime decodable words in text that would support instruction in using orthographic patterns to identify words. Still, when 84.1% of the books in the sample contain 27.4% (the mean plus one standard deviation) or fewer onset-rime decodable words, and over half the books contain 14% or fewer, it would seem unreasonable to conclude that Reading Recovery books as a category support instruction in using onset and rimes to decode.

**Sentence-level demands.** The average Reading Recovery book in our sample had a mean of 9.6 words per sentence. On the Spache (1953) readability formula, this mean sentence length is associated with a predicted grade level of at least 2.2. This mean, converted to sentences per 100 words, falls on the central line of the Fry Readability Graph (Fry, 1977) at the cusp between second- and third-grade level. Both these estimates are somewhat high given that Reading Recovery books have a median grade level of 1.5 or less.

The mean T-unit length of the 80 books was 6.4 words and 7.0 morphemes. To our knowledge no norms exist for interpreting the T-unit length of comprehensible input, whether written or oral, at various ages or grades. However, Loban (1976) computed average number of words per communication unit (equivalent to Hunt’s T-unit) in expressive oral language across grades 1–12 and in expressive written language across grades 3–12. Using those averages, the mean T-unit length of 6.4 words in Reading Recovery books is noticeably less than first graders’ expressive oral language (7.9) and third graders’ expressive written language (7.7). Because the syntactic complexity of young children’s expressive language is generally less than the syntactic complexity of the receptive language they can comprehend (Paul, 2001), the sentence-level demands of Reading Recovery books appear to be very moderate.

When contrasting average sentence length and T-unit length in words, it appears that the estimate of sentence-level demands employed by readability formulas such as the Spache may be quite mislead-
ing, at least in books written for first graders to read. Apparently, Reading Recovery books have longer sentences than the books of the 1950s, but they still maintain a moderate syntactic complexity for young children. Specifically, Reading Recovery books probably have more compound sentences without having more syntactic complexity than a readability formula such as the Spache would predict.

**Discourse-level demands.** The average Reading Recovery book in our sample had 214.2 running words and 70.1 unique words. Children in the Reading Recovery program, and in regular classroom programs affiliated with it, are expected to be able to read about 90% of the running words in a text the first time they read it (Clay, 1993). Our findings suggest that a student reading a Reading Recovery book with that degree of challenge would probably encounter about 7 unique words that were new or challenging with about 3 occurrences each. In Reading Recovery, these encounters presumably provide the student with the desired number of opportunities to make oral reading errors or miscues that elicit coaching instruction while using the three main cueing systems to solve problems. What about classrooms where the teacher emphasizes word recognition or decoding instruction rather than the three main cueing systems? Using a 90% criterion of oral reading accuracy for determining which level of books a student should be given to read would yield a maximum of 21 opportunities for students to apply what they are being taught; a 95% criterion would yield a student about half that many opportunities.

**Conclusion.** From our analysis of their word-, sentence-, and discourse-level demands, we conclude that Reading Recovery books, as a category of texts for early reading instruction, provide some support for an instructional emphasis on the recognition of high-frequency words, but inadequate support for an instructional emphasis on decoding words using onsets and rimes. The strength of the Reading Recovery books as a category of reading instructional texts appears to be primarily at the sentence level. If Reading Recovery books in general contained a small increase in percentage of high-frequency words and a somewhat higher type-token ratio, their average sentence-level demands could be expected to provide much better support for students in self-monitoring their word recognition than traditional instructional text based on readability formulas ever did.

Likewise, the moderate syntactic complexity of Reading Recovery books, achieved while avoiding short unnatural sentences, would seem to provide ideal support for students in self-monitoring their on-set-rime decoding as well. Unfortunately, Reading Recovery books as a category of texts comprised of useful onsets and rimes for these sentence-level advantages to come into play or for us to interpret the potential usefulness of the average number of decoding opportunities afforded by this category of books at the discourse level.

**Instructional supportiveness of the leveling of Reading Recovery books for instruction emphasizing word recognition or decoding**

The results of the various correlational analyses between Reading Recovery level and the measures of print-based demands at three levels of text structure were interpreted to characterize the instructional supportiveness for word recognition or decoding of how Reading Recovery books are leveled.

**Word-level demands.** The Reading Recovery books in our sample did not become more challenging as book level increased along any of the word-level dimensions assessed by our nine measures.

**Sentence-level demands.** Sentence length, whether in words or morphemes, did not significantly increase at higher Reading Recovery levels. This finding supports the contention that readability formulas using sentence length as a variable cannot predict the level of Reading Recovery books (Peterson, 1991). However, T-unit length in both words and morphemes did significantly and positively correlate with Reading Recovery level, indicating that increased syntactic demands are associated with Reading Recovery leveling. Although not related to our research question, this finding lends some support to the claim that Reading Recovery leveling supports instruction in the three main cueing systems. Because the syntax of the sentences in a text is considered one of the main cueing systems (Clay, 1993), a sequence of texts that supports instruction in the three main cueing systems should increase in its syntactic demands as text level increases.

Pertinent to our research question, the increase in syntactic demands at higher levels of text is consistent with what would support instructional emphases on word recognition or decoding. Indeed, many advocates of teaching high-frequency words or decoding systematically also advocate teaching students to use context to self-monitor their word iden-
Discourse-level demands. All five measures of book length were significantly and positively correlated with Reading Recovery level, indicating that discourse-level demands increase as assigned level increases. Hiebert (1999) remarked that, “Within current schemes of text selection for beginning readers, the only concession to word difficulty is attention to text length” (p. 560). Our findings replicate hers. The best overall predictor of Reading Recovery level was the number of unique words (types) in the book. This single print-based measure accounted for 47% of the variance in Reading Recovery level across the 80 books in our sample. Although not related to our research question, this finding lends credence to the claim that Reading Recovery leveling supports instruction in the three main cueing systems. Because children in Reading Recovery are taught to use the three main cueing systems to solve problems when they arise (Hicks & Villaume, 2000), we would expect harder books in a sequence supporting such instruction to have more problems to solve (more unique words to figure out) as children become more proficient.

Relevant to our research question, the increase in discourse-level demands at higher levels of text is consistent with what would support instructional emphases on word recognition or decoding. That is, the better students can recognize or decode words, the longer they should be able to identify and self-monitor words before becoming frustrated or fatigued. Moreover, the fact that number of unique words (types) rather than number of running words (tokens) was the better predictor of Reading Recovery level suggests that Reading Recovery leveling would support decoding instruction better than word-recognition instruction if the word-level demands were present to support such instruction.

Conclusion. The increase in sentence- and discourse-level demands as Reading Recovery level increases would support an instructional emphasis on word recognition or onset-rime decoding. Unfortunately, the Reading Recovery books in our sample did not become more challenging as book level increased along any of the nine word-level dimensions we measured. If our sample of Reading Recovery books was representative, this finding means that the way Reading Recovery books are leveled does not support instruction in recognizing high-frequency words or decoding words consisting of high-utility onsets and rimes.

Implications

The instructional supportiveness of Reading Recovery books for instruction emphasizing word recognition or decoding

The sentences in Reading Recovery books avoid the artificial shortness associated with text written to achieve a low readability formula score, yet they still present only moderate syntactic complexity that gradually increases as book level increases. These sentence characteristics are what one would expect to find from materials designed or selected to have natural language patterns that support the use of syntax as one main cueing system for solving problems while reading.

The best print-based predictor of Reading Recovery level is the number of unique words in a book. This discourse characteristic is what one would expect from books designed or selected and sequenced to provide students with an optimal number of opportunities (unfamiliar words) at each point of their development to use the three main cueing systems to solve problems (figure out the words).

However, books originally leveled for use in Reading Recovery are now also widely used in regular and special classrooms having no affiliation with Reading Recovery. This led us to pose the research question for this study: Do books leveled for use in Reading Recovery support other reading instructional emphases besides the use of the three main cueing systems?

Reading Recovery books as a category of instructional texts for early reading average fewer opportunities to recognize high-frequency words than would be expected from first-grade materials with the goal of supporting word-study instruction on such words. The type-token ratio of Reading Recovery books in general is lower than one would expect in texts that support instruction in word recognition. They also average fewer opportunities to decode words consisting of high-utility onsets and rimes than one would expect from books chosen or constructed to support such instruction. Teachers who attempt to teach their students to recognize high-frequency words without relying on context, or
to decode words comprised of common onsets and rimes without relying on context, should find texts more supportive of that instruction than Reading Recovery books.

Moreover, the way Reading Recovery books are leveled provides no support for instruction in recognizing words by their orthography or decoding them by their phonology. We recommend that advocates of word recognition or decoding instruction either seek other kinds of materials or select and rellevel a subset of Reading Recovery books that will provide increasing word-level demands as the assigned levels of the books increase.

Reading Recovery lessons incorporate both study of high-frequency words and phonics instruction that includes using onset-rime patterns. Yet the books they select for their program provide little support for these two instructional components, and the way they level the books provides none at all. The Reading Recovery Council of North America may want to consider revising their leveling approach to also take into consideration the occurrence of high-frequency and onset-rime decodable words. If it is unreasonable to expect a single category of texts to support word recognition and decoding instruction as well as instruction in the three main cueing systems, Reading Recovery may want to consider adding two additional lists of single-criteria texts for teachers to use that support their high-frequency word and onset-rime decoding instruction, respectively.

Books leveled for use in Reading Recovery and reading assessment

The results of this study warn us about the practice of schools and school districts where primary-grade children's oral reading is assessed using Reading Recovery books or passages. Because those assessments typically yield a reading level based on the Reading Recovery level of the books or passages used, they may not be valid for students whose teachers have emphasized word recognition or decoding rather than or more than the use of the three main cueing systems.

Reading educators may be able to select and rellevel subsets of Reading Recovery books that support the development of both word recognition and decoding

While the designation of levels [in this chapter] is a useful way to indicate roughly a progression of difficulty for books used during Reading Recovery lessons, such a system may not be suitable for other settings. (Peterson, 1991, p. 128)

This study found that books leveled for use in Reading Recovery do not consistently increase in word-level demands as their levels increase. As we have reasoned elsewhere (Cunningham et al., 2004), there are so many attractive and interesting books leveled for use in Reading Recovery that there is reason for optimism about the possibility of finding subsets of these books with ample percentages of high-frequency and onset-rime decodable words that gradually decrease as book difficulty increases. Perhaps such a new selection and leveling scheme could permit many of these books to continue to be used by teachers whose instruction emphasizes word recognition or decoding more than the use of three cueing systems.

Implications for further research

Because so little is known about the contribution of texts to early reading instruction, much research of various kinds is needed. At the least, other categories of reading instructional texts besides Reading Recovery books could be examined for evidence that they support particular instructional emphases. For example, the phonetically decodable texts currently mandated in some states could be analyzed to determine the extent to which they also support the teaching of high-frequency word recognition and decoding by onset-rime patterns. Also, given that certain kinds of texts have been mandated without evidence of their effectiveness, intervention studies of the instructional value of different categories of texts are needed. Taking an approach similar to the one in this study to determine the instructional supportiveness of the texts used in any intervention will increase the likelihood that the results of such studies will be interpretable.

The results of our study raise the question of how important leveling is as a separate factor from category in supporting early reading instruction. It has long been assumed in reading education that texts have levels and that texts should be sequenced by level when teaching reading. This assumption is the basis for readability and text leveling, ability and leveled grouping for reading instruction, and assessment of students’ reading instructional levels with informal reading inventories or benchmark books. We share that assumption and have employed it in this study. However, what if the levels assigned to books aren’t as important as their category? That is, what if the characteristics of the books students read,
more than the order in which they read them, shapes their reading strategies and supports the reading instruction they receive? We could find no previous research that raised or investigated this question, but future research could be designed to distinguish the effects on reading achievement of category from leveling in the materials used to teach reading in the primary grades. Until such research is done, however, it seems prudent to follow the consensus of professional opinion that books for early reading instruction should be leveled, and leveled along the curricular dimensions of the instructional emphasis the books are expected to support.

JAMES W. CUNNINGHAM is professor emeritus of literacy education at the University of North Carolina at Chapel Hill where he taught undergraduate and graduate courses in reading and writing education for 28 years. His continuing professional interests include improving elementary classroom writing instruction and investigating first graders’ sensitivity to orthographic patterns. He can be contacted at 811 Leigh Drive, Gibsonville, NC 27249-2734, USA, or by e-mail at jwcunnin@email.unc.edu.

STEPHANIE A. SPADORCIA is an assistant professor of language and literacy at Lesley University in Cambridge, Massachusetts. She teaches courses in the areas of struggling readers and writers, assessment, literacy for children with intensive special needs, and language and literacy development. Her research and writing focus on improved literacy instruction for struggling readers and writers. She can be contacted at 29 Everett St., Lesley University, Cambridge, MA 02138, USA, or by e-mail at sspadorc@mail.lesley.edu.

KAREN A. ERICKSON is the director of the center for literacy and disability studies at the University of North Carolina at Chapel Hill. Her research focuses on literacy assessment and instruction for children with significant disabilities including those who use augmentative and alternative communication. She is the 2004 recipient of the National Down Syndrome Congress Educator Award and the International Society for Augmentative and Alternative Communication Distinguished Literacy Lectureship Award. She can be contacted at CB# 7335, UNC-Chapel Hill, Chapel Hill, NC 27599-7335, USA, or by e-mail at Karen_Erickson@med.unc.edu.

DAVID A. KOPPENHAVER is an associate professor in the Language, Reading, and Exceptionalities Department at Appalachian State University in Boone, North Carolina. His research focuses on literacy instruction and assessment of children and adults with developmental disabilities, including cerebral palsy, autism, and intellectual disabilities. He can be contacted at Appalachian State University, LRE Department, 124 Edwin Duncan Hall, Boone, NC 28608, USA, or by e-mail at koppenha@apstate.edu.

JANET M. STURM is associate professor in the Department of Communication Disorders, Health Professions Building, 2167, Central Michigan University, Mount Pleasant, MI 48859, USA, or by e-mail at sturm1@cmich.edu.

DAVID E. YODER is professor emeritus of speech and hearing sciences at the University of North Carolina at Chapel Hill. His current research is focused on the literacy needs of persons with severe speech and physical impairments and assistive technology for adults with disabilities. Yoder is a member of the research team of the Center for Literacy and Disabilities Studies at the University of North Carolina, which is directed by Karen A. Erickson. The team is working to develop an alternative reading comprehension assessment battery for students with severe speech and physical impairments. He can be contacted at CB# 7335, TR# 46, UNC-Chapel Hill, Chapel Hill, NC 27599-7335, USA, or by e-mail at dyoder@med.unc.edu.

REFERENCES


HUNT, K.W. (1964). Differences in grammatical structures written at three grade levels, the structures to be analyzed by transformational methods (Cooperative Research Project # 1998). Tallahassee, FL: Florida State University.


Received December 15, 2003
Final revision received August 23, 2004
Accepted November 5, 2004