Critical Word Factor in Texts for Beginning Readers

ELFRIEDA H. HIEBERT
University of California, Berkeley

CHARLES W. FISHER
University of Michigan

ABSTRACT **The *Critical Word Factor*, based on word recognition demands of texts, is a measure of text difficulty designed specifically for texts used by beginning readers. The measure is a function of the number of new, unique words per 100 running words of text that fall outside a designated curriculum. The authors investigated the validity of the Critical Word Factor from data on beginning readers' speed, accuracy, and comprehension after students read texts with different numbers of critical words. Analyses of variance indicated significant main effects for the Critical Word Factor on reading speed, accuracy, and comprehension. Mean differences on the 3 variables were in the predicted direction; results for speed and accuracy were stronger than were those for comprehension. Additional analyses showed that words predicted by the model to be hard were hard, and those predicted to be easy were easy.**

Keywords: beginning readers' speed, accuracy, and comprehension; Critical Word Factor, student reading outcomes

Conceptions of reading draw typically on characteristics of texts, characteristics of readers, and the myriad interactions between these two factors (Anderson & Pearson, 1984). School programs designed for teaching reading give more or less weight to each of these basic domains. The design of new reading programs, analysis of existing programs, and their implementation in classrooms has led to the invention of a variety of indexes for characterizing the level, difficulty, or accessibility of texts. We introduce an index called the *Critical Word Factor* (CWF), specifically for texts used by beginning readers. The index is not a blanket measure for reading texts but is intended to describe texts at the stage in which children, regardless of their age, acquire basic reading proficiency. We describe the CWF in terms of word-recognition demands of texts and comment on the relationship between the CWF and several current indexes. We also present data on beginning readers' speed, accuracy, and comprehension after reading texts that differ in CWF.

Recent Changes in Text Characteristics

In the past several decades, the characteristics of texts for beginning (as well as more proficient) readers have undergone a series of major shifts. In 1987, the California State Board of Education formally adopted authentic literature (i.e., trade books) for elementary students rather than texts with controlled vocabulary that had been used for the preceding decades. This decision heralded rapid change in the texts used by beginning readers in California. In 1990, the Texas State Board of Education made a similar policy decision. Because California and Texas represented large market segments for text publishers, many other states were influenced indirectly by decisions made in Sacramento and Austin. Primarily because of that influence, many American school systems replaced existing texts for beginning readers with authentic children's literature.

However, in the succeeding textbook adoption cycle, California and Texas opted for decodable texts (California English/Language Arts Committee, 1999; Texas Education Agency, 1997); their decisions led to another round of replacement of texts for beginning readers in a substantial number of states. Those shifts, from the types of texts that prevailed before 1987, to authentic literature in the 1990s, to the current decodable texts, represented dramatic swings in the features of texts used by beginning readers.

Changes have been so large that one could perceive the cognitive tasks required of beginning readers from one decade to the next as differences in kind rather than in degree. Consider, for example, Scott Foresman's reading textbook program, which has the longest record of publishing in the field. From 1962 to 2000, the number of unique or different words (per text) increased from 18 to 187 for the first 10 texts in the program (Hiebert, 2005). That 10-fold increase in the number of unique words to be read represents a sea change in developmental expectations for beginning first-grade students. One would have difficulty finding an adequate research base for such an ambitious rate of word introduction for beginning first-grade readers, particularly those whose success depends on the quality of

Address correspondence to Elfrieda H. Hiebert, 106 Phelan Court, Santa Cruz, CA 95060. (E-mail: hiebert@berkeley.edu)

their school experiences. Moreover, a substantial percentage of American fourth-grade students are not attaining national standards (Donahue, Finnegan, Lutkus, Allen, & Campbell, 2001), a pattern that apparently begins in first grade (Juel, 1988).

Those changes in text features were tantamount to a series of national interventions in a critical aspect of beginning reading instruction. Surprisingly, the theory and research underlying these interventions were not extensive. Although an extensive literature exists on beginning reading, there are remarkably few studies on the effects of text features on students' reading speed, accuracy, and comprehension, and even fewer studies focus on beginning readers. Understanding how various features of text support or hinder the progress of beginning readers appears to be understudied (Hiebert & Mesmer, 2005). We undertook the current work to partially redress the situation.

Existing Frameworks of Text Difficulty

For much of the 20th century, the dominant work on text difficulty centered on the notion of readability. Various text features like *semantic complexity* (e.g., presence of words on a designated list; number of syllables per word) and *syntactic complexity* (e.g., number of words per sentence) were combined in various ways to create readability indexes. Generally speaking, readability formulas weighted the characteristics of texts much more than either the cognitive processes of the reader or the interaction of these two entities. Although the formulas provided useful results for relatively advanced readers, results for the earliest stages of reading acquisition were less satisfactory. An example is the popular text *Little Bear* (Minarik, 1957) that appears at the end of two current textbook programs (Farr et al., 2001; Flood et al., 2001). When readability formulas are applied, *Little Bear* yields readabilities that range from the middle of first grade to the middle of third grade (Micro Power & Light, 1999).

Problems with readability formulas were not limited to instability at the early stages of reading. Beginning in the 1980s, the application of cognitive science perspectives to understanding reading acquisition showed that strict compliance to readability formulas could have negative consequences for comprehension (Bruce, 1984; Green & Davison, 1988). When experimenters modified texts to lower readability indexes either by substituting high-frequency words for less frequent words or by shortening sentences, student comprehension declined (Green & Davison, 1988).

Critiques of readability, among other factors, eventually led to the mandates by California and Texas (described earlier) for authentic literature rather than texts that had been manipulated to comply with readability formulas. Because cognitive scientists had not described text features that support children at the earliest stages of reading at the time the mandates were initiated, publishers had very little guidance when identifying authentic literature to be used with beginning first-grade students.

Hoffman et al. (1994) analyzed the texts adopted by the Texas State Board of Education in 1993. Their analysis showed a preponderance of the "predictable" text genre in first-grade textbooks. That genre consists of repetitions of rhythmic and rhyming words, phrases, or sentences, allowing beginning readers to predict many words (Rhodes, 1979). Research on predictable text was almost nonexistent at that point. Furthermore, when Johnston (2000) reviewed the handful of existing studies, the results showed that overuse of predictable texts did not develop attention to word features, particularly among children at the early stages of word recognition.

As educators witnessed the effects of a diet of predictable texts during the beginning reading phase on subsequent development of reading proficiency, large numbers of reading educators sought alternative types of texts. The Texas Education Agency (1997) responded by prescribing percentages of decodable words to be included in beginning reading textbooks. In their 1999 textbook guidelines, the California Department of Education implemented a similar prescription.

Scholarship on the appropriateness of text features for beginning readers has not moved as rapidly as have state and district policies. However, Hiebert and Mesmer (2005) identified four relatively new text difficulty schemes: (a) lexiles (Smith, Stenner, Horabin, & Smith, 1989), (b) leveled texts (Fountas & Pinnell, 1999), (c) predictability and decodability (Hoffman, Roser, Patterson, Salas, & Pennington, 2001), and (d) potential for accuracy (Beck & McCaslin, 1978; Stein, Johnson, & Gutlohn, 1999). To illustrate those schemes, consider the difficulty ratings they generate when applied to *Little Bear*. The text is described by a lexile of 370, at the top of the 200–370 span that is recommended for Grade 1. The same text is assigned a guided reading level of J (Fountas & Pinnell, 1999). Level J texts are the 8th of 13 levels designated as the first-grade range and are distinguished from prior first-grade levels by their length and inclusion of dialogue.

On the basis of criteria described by Hoffman et al. (2001), *Little Bear* would receive a predictability rating of 4 (somewhat predictable) and a decodability rating of 3 (moderately decodable text); both the predictability and decodability scales have a range from 1–5. As expected, Hoffman et al. (2001) found that better readers were able to read less predictable and less decodable texts. However, those ratings are not explicitly grounded in a framework describing the roles of the variables at different points in children's reading development.

The potential for accuracy scheme expands the scope of the earlier indexes by considering not only text features but also what is expected to have been taught before a reader encounters a given text. A student is considered to have potential for accuracy on a word if (a) the word is a high-frequency word that already has been taught or (b) if the word is decodable and each of the decodable elements in the word has been previously taught. In the typical implementation of this measure, information about instruction comes from the teacher's manual rather than from actual instruction. Potential for accuracy was assessed in that way for all of the first-grade reading programs submitted for adoption in Texas. As part of the adoption process, the Texas State Board of Education required that for placement on the conforming list, programs had to include texts that were at least 80% decodable (Stein, Johnson, Boutry, & Borleson, 2000). The potential for accuracy scheme assumes that every teacher is fully and accurately implementing the teacher's manual. However, for a given student, it is not clear whether the relevant instruction occurred, when it occurred, or in what context it occurred. Furthermore, a given student may have required several exposures to content to apply it independently. We used *Little Bear* in this section of our article to illustrate difficulty indexes. Because the potential for accuracy assigned to *Little Bear* depends on the teacher's manual for the program in which the text is embedded, we are unable to provide a unique numerical index.

Each of these four approaches to assessing text difficulty weighs text characteristics, reader characteristics, and their interactions, although in various combinations and degrees. None of the indexes is particularly transparent to interpret, although each has merit. It is difficult for one to incorporate reader characteristics and the interactions of text and reader characteristics in a satisfactory manner. The potential for accuracy goes farthest in that direction by distinguishing between text features for which the reader can be expected to have been exposed and those for which the reader has not been exposed. Where information that has been taught is well known, potential for accuracy has the potential to be useful.

Most current approaches to text difficulty, including lexiles and leveled books, have reported little or no data on children's reading performances as a function of one or more of the indexes. The design of textbooks and instruction for beginning reading could benefit from more and better indexes of text difficulty, especially indexes with empirical support.

CWF

CWF is an index of a text's difficulty for learners who are working on a specific set of reading tasks. Like other indexes, the CWF depends on the configuration of high frequency and decodable words in a text, but unlike other indexes, the CWF also depends on the curricular domain within which an indexed text is to be used. The value of the CWF for a given text changes as the reading domain of interest expands or contracts. That allows the CWF, at least in principle, to be tailored for different groups of readers. The importance of specifying the reading domain is illustrated by an analysis of typical tasks encountered by beginning readers.

The following two sentences, taken at equivalent points of two different beginning reading programs, illustrate differential task demands for beginning readers:

Example 1: I can hop, run, and dig.
Example 2: I found my old, orange tiger.

Although each sentence has the same number of words, the sentences differ in the linguistic knowledge demanded of beginning readers. In the first sentence, knowledge of simple vowel patterns (i.e., where a single grapheme represents a single phoneme, as in *go* or *cat*) is adequate for recognition of all of the words.

Application of the same knowledge to the second sentence leaves several words unrecognized. The inclusion of words such as *old* and *found* in the second example introduces complex vowel patterns. Those two words are among the 200 most frequently used words in English texts. In addition, other common words have similar vowel patterns (e.g., told, cold; round, ground). However, those vowel patterns fall into the complex vowel category. Also, by including the word *orange*, beginning readers must differentiate three different sounds associated with the grapheme *o* in one six-word sentence. The early introduction of complex vowel patterns to beginning readers has been common practice in American reading instruction (Gates & Russell, 1938–1939; Gray, Monroe, Artley, Arbuthnot, & Gray, 1956). Historical precedents notwithstanding, the cognitive demands of reading the two sentences (i.e., Examples 1 and 2) are remarkably different.

To become a fluent reader of English, beginning readers must become proficient at several core tasks. They must learn a set of relationships between letters and sounds (Adams, 1990; National Institute of Child Health and Human Development, 2000) and learn which relationships to use in a large variety of contexts. That is, while learning the standard letter–sound relationships, beginning readers attempt to develop a set for diversity (Gibson & Levin, 1975). Those tasks are more or less forced on beginning readers simultaneously because so many high-frequency words have irregular vowels. Of the 25 most frequent words in written English, half have irregular vowel patterns. Because these 25 words account for one third of the total number of words in texts (Carroll, Davies, & Richman, 1971), children must recognize them quickly to become successful readers.

Graphic illustrations in books can permit children to make one-to-one matches between objects and words, making the immediate task of naming objects easier. However, the effect on later recognition of words occurring without illustrations appears to be negligible, especially for children at the earliest stages of reading (Johnston, 2000).

Our brief and incomplete analysis demonstrates the high level of cognitive processing demands on beginning readers. Words that are not easily decodable or highly frequent increase the already high level of cognitive demand. The greater the number of such words in a text or series of texts, the higher the cognitive processing demands on beginning readers. As processing demands increase, rate and accuracy of word recognition likely decrease. If rate and accuracy of word recognition decrease, comprehension may also decrease.

The CWF is a function of the task demands for recognizing words in primary-level texts. Specifically, the CWF indicates the average number of *hard* or *critical words* (i.e., those that fall outside specified high-frequency and phonics curricula) in 100 running words of text. The CWF assumes the existence of an underlying curriculum related to word recognition. For example, if the end of first-grade curriculum is the 300 most frequently used words and all vowel patterns in single-syllable words except for diphthongs and variants, the CWF for *Little Bear* is 2. Of every 100 words of text, readers who are proficient in recognizing the 300 most frequently used words and decoding most single-syllable words will encounter 2 hard words. If the curriculum were designated as the 100 most frequent words used and phonetically regular words with simple vowel patterns, then the CWF would be 8. In other words, one would expect that 8 of every 100 words would be difficult.

Reading Performance on Texts Differing in CWF

Examining first-grade students' reading acquisition when taught with various types of textbooks, Juel and Roper/Schneider (1985) identified the first term of first grade as the time when students were most influenced by text features. Consequently, we conducted the present study at the end of the first trimester of first grade.

We used existing texts to study the effects of the CWF. We used texts from "little book" programs that are used widely in American classrooms. The little books are typically advertised as leveled, according to the guided reading criteria of Fountas and Pinnell (1999). The selection of particular texts is described more fully in the Method section.

We examined the reading speed, accuracy, and comprehension of first-grade students on texts differing in CWF. Thirty-six students read four texts (two low-CWF and two high-CWF texts) in a repeated-measures design. All participants progressed within the expected or typical range for first grade. We collected data during 2 weeks at the end of the first trimester of Grade 1. The primary research question was, "Do students read low-CWF texts faster, more accurately, and with higher comprehension than they do high-CWF texts?"

Method

Sample

Participants attended two schools in a medium-sized western U.S. city. In both schools, approximately 40% of students received free or reduced-price lunches. Approximately 25% of the students in each school were English language learners (ELLs), and this percentage was near that of the state's average (Donahue et al., 2001). We selected children from six first-grade classrooms—four in one school and two in the other school.

Pilot testing revealed that some students in the participating schools either named known letters in the text or pretended to read a story. Those students were performing at the earliest of Sulzby's (1985) book-reading stages. Because those responses indicated little about the relative difficulty of the texts, we used a word-recognition task as a screening measure.

The screening measure began with a list of 10 high-frequency words drawn from the 25 most frequently used words (Carroll et al., 1971). The sample included only those children who read at least 5 of the first 10 words correctly. Children who attained that level continued with the task until they failed to read 6 consecutive words on 7 subsequent 10-word lists representing progressively less frequently used words on the Carroll et al. list. Of the 36 students who correctly read 5 of the first 10 words, scores ranged from 5 through 79, with a mean of 37 and a median of 31. Five students scored higher than 70, and 4 students scored lower than 10. Of the 36 students who were successful on the screening measure, 15 were boys and 21 were girls. That group included 5 students (2 boys and 3 girls) who were ELLs. The sample represented a broad range of reading proficiency but less than the full range of proficiency found in the two schools.

Materials

We used four texts representing two levels of CWF (two low-CWF and two high-CWF texts). We selected all of the texts from little book programs in common use; we used existing texts because of their face validity for educators.

We chose texts from a pool of texts identified in a previous study (Hiebert, 2001). In that study, we calculated the CWF for 10 little books at each of seven steps within six different little book programs (i.e., a stratified sample of 420 little books). The six little book programs included Harcourt Collections (Farr et al., 2001), Rigby PM Plus Program (Rigby Education, 2000), Ready Readers (Modern Curriculum Press, 1997), Sunshine Reading Program (Wright Group, 1996), Waterford Early Reading Program (Waterford Institute, 2000), and Open Court decodables (Adams et al., 2000). Except for the Open Court decodables, publishers provided information on the levels of texts according to the Reading Recovery scale (Peterson, 1991) or the guided-reading levels (Fountas & Pinnell, 1999). For the period under study (the end of the first trimester of first grade), the recommended levels were 3–4 (Peterson, 1991) or levels C–D (Fountas & Pinnell, 1999).

We analyzed 60 little books (10 from each program with first-trimester, Grade 1 text levels) to determine how many words per 100 did not fall within the specified curriculum (the 100 most frequently used words and vowel patterns in single-syllable words through CCVC and CVCC patterns). The texts from four of the programs (Harcourt, Rigby, Waterford, and Wright) had an average CWF of 22; that is, 22 out of 100 running words of text fell outside the specified curriculum. Within this set of 40 texts, the CWF ranged from 6.6 to 47.9, with a median of 17. Texts from those programs with a CWF between the median (17) and the mean (22) constituted the pool from which we selected the high-CWF texts.

Texts of the other two programs (Ready Readers, Open Court decodables) had an average CWF of 2.5. The range was 0 to 10; median equaled 2. Forty percent of the texts in the two programs had a CWF of 0 (i.e., the texts had no words beyond the specified curriculum). We selected the low-CWF texts from that pool.

From the two pools, we selected four texts (two high-CWF and two low-CWF texts). In the final selections, we considered two additional factors. We chose texts with sufficient content to allow assessment of comprehension from the available texts. The meaning of the text had to be apparent from the words alone. Using those criteria, we selected four texts: (a) one text from Sunshine (CWF equal to 21), (b) one text from Rigby PM (CWF equal to 20), and (c) two texts from Ready Readers (CWF equal to 0).

We made two types of minor modification to the selected texts. First, to ensure that each text had an approximately equivalent number of words (within a range of 53 to 56 words), we deleted a few words from some texts. Deleted words did not change the meaning of the texts. Second, we modified the titles of the texts to make them similar for low- and high-CWF texts. Each title contained a critical phrase from the text. One title of each pair was a two-word phrase and one was a three-word phrase. Examples of a low- and a high-CWF text are shown in Table 1. We summarized the characteristics of the four texts in Table 2.

The students received the texts with a single illustration on each booklet cover along with the title (the titles were read to students). We chose each illustration from a clipart library; they were related generically to the text. With the exception of the cover, the texts did not have illustrations. Therefore, we attempted to assess reading performance on the basis of the text, independent of clues that might have been available from illustrations.

TABLE 1. Examples of Low- and High-Critical Word Factor (CWF) Texts

Low-CWF text	High-CWF text
Hop! Hop! Hop!	*My Book*
Hop, hop, hop on the bed.	Where is my new, red book? My book is not in the bed.
"Stop! Stop! Stop!" said Dad. "Not on the bed."	I found my old, gray elephant.
Hop, hop, hop in the bath.	My book is not over the bed.
"Stop! Stop! Stop!" said Dad. "Not in the bath."	I found my old, brown monkey.
Hop, hop, hop on the mat.	My book is not under the bed.
"Yes, yes, yes," said Dad. "Hop, hop, hop on the mat."	I found my old, orange tiger.
Hop, hop, hop.	I found my new, red book!

TABLE 2. Characteristics of Texts

Text	Total words	Unique words: Total	Unique words: Per 100 words	High-frequency rating[a]	Decodability rating[b]	Hard words per text
Low 1	53	12	23	5.3	2.5	0
Low 2	54	11	20	5.7	2.3	0
High 1	53	21	40	4.3	5.0	11
High 2	56	22	39	4.3	5.1	10

[a]Higher values represent higher frequency words. [b]Higher numbers represent less decodable words.

Procedure

Each student participated in two short sessions over a 2-day period. We conducted the one-on-one sessions in a convenient area in the school but outside the child's classroom. The first session began with the screening measure. If the students did not read at least 5 of the first 10 words correctly, the experimenter thanked them, and they returned to the classroom. Students who read correctly 5 of the first 10 screening measure items continued until they missed 6 consecutive items or until they reached the 80th word. We recorded responses for each word as correct or incorrect.

When students completed the word-recognition task, we gave them one of the four experimental texts. Each student read four passages over the two sessions. During the first session, students read one low- and one high-CWF passage. In the second session, students read the other two passages (one high and one low CWF), with the low–high ordering reversed. Because there were two examples of low- (and high-) CWF passages, reading order effects were still possible. To guard against such effects, we listed the eight possible orders and used them sequentially during data collection, ensuring that reading orders were balanced.

Two trained data collectors gathered the data. One researcher worked with a student during the first session, and the second researcher worked with that student during the second session. Half of the students read with Researcher A for the first session and half read with Researcher B, balancing the effect of data collector over the two sessions. One of the data collectors was male and one female. To prevent possible gender imbalances, the male data collector conducted the first reading session with half of the boys and half of the girls in the sample.

Text presentation began with the researcher reading the title to the students, then asking them to continue reading the text. At that point, the researcher began timing the length of time that it took the student to read the text. As the students read, the researcher recorded their miscues, focusing on omissions, substitutions, and insertions. The researcher also recorded the start and stop times of students' reading of each text. Following the completion of the students' reading, the researcher asked the students, "Can you tell me what the story was about?" followed by the prompt, "Can you tell me anything else that happened?" The researcher wrote down students' responses verbatim. In the second session, the student met with the second researcher and read two additional texts.

Scoring

Trained raters assessed students' readings of each of the four texts for speed and accuracy and scored their responses to the comprehension prompts. Each child read four passages, so there were four occasions for each of the three measures of speed, accuracy, and comprehension.

The *speed variable* was the number of words read divided by the time elapsed for reading the passage. The metric of the speed variable was words per minute. We did not consider students' reading errors for this variable.

The *accuracy variable* measured the extent to which a child's reading of a passage was error free. We obtained the score for a given passage by taking the number of words wrong plus one error for every insertion (regardless of number of words in that insertion), subtracting this quantity from the total number of words read in the passage, dividing the result by number of words read, and multiplying by 100. Accuracy was equal to $N\text{-}E/N \times 100$ (where N = total number of words read in the passage and E = number of errors + number of insertions). The metric of the accuracy variable was words correct per hundred words.

We rated comprehension from students' responses to the question, "Can you tell me what the story was about?" and a follow-up prompt asked immediately after their reading of the story. The study texts were short and used a predictable text structure that did not include all of the elements associated with story grammars. These features made it difficult to use a retelling protocol such as that developed by Morrow (1986) for evaluating student responses. We used a 5-point rating scale modeled after that used by the National Assessment of Educational Progress (Donahue et al., 2001) for scoring responses to text segments. A score of 0 represented no evidence of text comprehension, as in responses such as, "I don't know." A score of 1 indicated minimal evidence of text comprehension, as evidenced by use of the title or the illustration on the cover. We awarded a score of 2 for evidence of a few concepts or actions from the text beyond the title, whereas a score of 3 indicated evidence of the theme of the text, including a representation of most of the actions or concepts in the text. Students received a score of 4 for responses that indicated full comprehension of the text in which they used at least some details from the text to elaborate the theme.

Two trained raters assessed each response independently. Subsequently, the raters examined differences, and if agreement was reached, they amended a rating. Amendments were never more than 1 category. When agreement could not be reached, they did not change the ratings. The average correlation between raters was .82 for the original ratings and .92 for the amended ratings. For analysis, we averaged the comprehension scores of the two raters.

Results

We used analysis of variance (ANOVA) to assess the effects of texts with different CWFs on reading performance. Analyses of reading speed, reading accuracy, and reading comprehension are presented in turn. For each outcome variable, we examined the following comparisons: (a) means for the two low-CWF texts, (b) means for the two high-CWF texts, and (c) sum of the means for low-CWF texts to the sum of the means for the high-CWF texts. In each analysis, we expected no differences for the first two contrasts (i.e., no differences in performances on texts with the same CWF). For the third contrast, however, we expected average performance on the low-CWF texts to be better than average performance on the high-CWF texts (i.e., faster, more accurate, and with higher comprehension). We also ran analyses of covariance, with word recognition as the covariate. The pattern of results was highly similar for both sets of analyses. Because the more complex analysis did not change the pattern of results, we reported only the ANOVAs in this study. Finally, we compared errors on individual words with the model's prediction of hard or easy words.

Effect of CWF on Reading Speed

Table 3 shows the means and standard deviations for reading speed on each of the four texts; ANOVA results indicate a strong main effect of the CWF on reading speed, $F(3, 105) = 44$, $p < .001$. The results for the three contrasts were as expected. Contrasts for pairs of texts with the same CWF (i.e., Low Text 1 to Low Text 2) were not significant ($p > .10$), indicating no differences in reading speed. The contrast between the two pairs (i.e., low CWFs to high CWFs) was significant ($p < .01$).

Effect of CWF on Reading Accuracy

Table 3 includes the means and standard deviations for reading accuracy on each of the four texts. ANOVA results indicate a strong main effect of the CWF on reading accuracy, $F(3, 105) = 35.5$, $p < .001$. The first contrast (pair of low-CWF texts) and third contrasts (low-CWF texts to high-CWF texts) followed the expected pattern. That is, reading accuracy did not differ between the two low-CWF texts but did differ significantly between the low- and high-CWF texts. The second contrast indicated that reading accuracy differed significantly for the two high-CWF texts ($p < .01$). Despite the unexpected difference between the two high-CWF texts, the means for accuracy on both high-CWF texts were lower than were those for the low-CWF texts. The mean for reading accuracy was the same for the two low-CWF texts, 86.7%, whereas the means for the high-CWF texts were 78.3% and 70.7%, respectively.

TABLE 3. Descriptive Statistics for Reading Speed, Accuracy, and Comprehension by Text

	Speed		Accuracy		Comprehension	
Text	*M*	*SD*	*M*	*SD*	*M*	*SD*
Low 1	47.0	24.7	86.7	14.3	1.8	0.9
Low 2	48.5	26.8	86.7	14.2	2.2	0.8
High 1	30.9	24.8	78.3	18.3	1.8	0.8
High 2	28.1	17.3	70.7	18.9	1.4	0.7

Note. $N = 36$.

Effect of CWF on Reading Comprehension

Table 3 shows the means and standard deviations for reading comprehension on each of the four texts. The ANOVA results indicate a significant main effect of the CWF on reading comprehension, $F(3, 105) = 10.9$, $p < .001$. The contrasts reveal that comprehension on like-CWF texts approached significance at $p = .05$. However, the third contrast (high-CWF texts compared with low-CWF texts) showed that reading comprehension in the low-CWF texts was significantly higher than in the high-CWF texts ($p < .01$). An examination of the means in Table 3 reveals that this difference is entirely between Low Text 2 ("See Me"; M = 2.2) and High Text 2 ("Up and Down"; M = 1.4); the other two texts made no contribution. Low Text 1 ("Hop, Hop, Hop") and High Text 1 ("My Book") had the same mean (1.8).

Reading Errors and Characteristics of Words

According to the model, words that are beyond the high-frequency list associated with the curriculum and not decodable within the specified limits of the curriculum are designated as *hard words*. The four texts in this study included 20 hard words. At least one third of the students made errors on 19 of the predicted hard words. The word book was the only exception. Although book was beyond the high-frequency list for the students and not decodable, it was a frequently used word in school conversations. Thus, it is not surprising that few children could not recognize it. Ninety-five percent of the words predicted to be hard by the model were hard.

The model predicted that the remaining 34 words in the texts would be easy for students to recognize. At least one third of the sample made errors on 5 of the words. Four of the five words (*bath*, *bog*, *dig*, and *spin*) were decodable, and the fifth word (*where*) was not decodable but was on the high-frequency list for the curriculum. Eighty-five percent of the words that the model predicted to be easy were easy.

We recorded the number of errors on each unique word in the texts and indexed them for decodability. A rating of 1 denoted an easily decodable word, and a rating of 8 denoted a very difficult word to decode. The correlation coefficient between errors per word and rating of decodability was .64. We also indexed unique words by frequency of occurrence; the correlation between errors per word and frequency of occurrence was .55.

Discussion

ANOVA indicated significant main effects for the CWF on reading speed, accuracy, and comprehension. The three variables were in the direction predicted by the model; results for speed and accuracy were stronger than were those for comprehension. In addition, words predicted by the model to be hard were hard, and those predicted to be easy were easy.

One should bear several caveats in mind when interpreting the results of the study. We selected texts with considerable care to reduce variation in characteristics other than CWF. However, the small number of texts used and the large number of potential variables on which texts can differ may have affected the results. We selected the sample of students by imposing a lower limit on a word-recognition task. The results are restricted to similar groups of students. The participants received the texts without illustrations, and, as a result, the findings are limited without additional experimentation.

Although it is beyond the scope of this study, the findings on reading speed may be pertinent to the design of texts for beginning readers. Analyses of sequential texts from widely adopted first-grade programs indicate that, whatever their philosophy, relatively large numbers of new words are introduced. Foorman, Francis, Davidson, Harm, and Griffin (2004) analyzed six first-grade programs, including programs on the 2000 Texas textbook adoption list. In four of the six programs, an average of 84 new words per week was introduced. Of the 84 new words, 70% consisted of difficult-to-decode words that appeared only a single time during a 6-week instructional block. The other two programs introduced new words at a somewhat lower rate. But even in those programs, over half of the new words appeared once in a 6-week instructional block.

Because the texts in this study with high CWFs had characteristics similar to the texts in Foorman et al.'s analysis, the slower reading of the high-CWF texts may be similar to the daily experiences of beginning readers in many American classrooms. The present study raises the concern that many beginning readers may spend large portions of daily reading time figuring out new words that do not recur in their texts. As a result, substantial numbers of children may not develop text-reading speeds as high as those that characterize proficient readers (Pinnell et al., 1995). Whereas more research is needed on that point, the CWF could inform the processes used by textbook publishers who develop the reading programs and ancillary materials used by the vast majority of America's beginning readers.

Comprehension may also be affected by texts with high percentages of difficult words (Shinn, Good, Knutson, Tilly, & Collins, 1992). Maintaining fluency at an appropriate level, especially at the early stages of learning to read, may be particularly important. The CWF may offer a new tool for research in this area.

In this study, the decodability of a given word was a good predictor of the number of reading errors on that word. That occurred even though the children were learning to read in classrooms in which explicit phonics was not a dominant instructional theme. Children appeared to develop generalizations about particular vowel and consonant patterns, even with limited or incidental instruction in phonics. However, words at the upper end of the designated decoding curriculum—short vowel words with consonant digraphs or blends in the initial or final positions—proved to be more difficult than predicted by the model.

We provided preliminary evidence that the CWF of a text can be used to predict important student reading outcomes. As a result, the CWF, focusing on the demands of word recognition, may be a more useful index of difficulty for the early stages of reading than are indexes with a considerably broader focus. For example, the four texts had similar guided-reading levels of C–D, but children still performed very differently on the low- and high-CWF texts. We presented texts without illustrations, which are a component of guided-reading levels. By partially removing the effects of illustrations, differences between CWF and guided-reading levels may have increased. However, reliance on illustration, especially when the number of difficult words is high, is unlikely to help children attend to word-level features (Samuels, 1970). We made no attempt to downplay the role of illustrations. Rather, it is important to deepen understanding of the various factors influencing children's beginning reading performances. We did not use illustrations with the experimental texts because we intended to assess children's word recognition without the distraction or facilitation that illustrations might have provided. However, each text cover did have an illustration to indicate the text's general theme.

The results of this study are sufficiently supportive of the CWF; therefore, we encourage four avenues of additional research. First, it is not clear which increment in the CWF will yield practical differences in readers' performances. We compared texts that differed in CWF by a relatively large amount. Therefore, we suggest that researchers examine smaller differences in CWF to explore the minimum difference in CWF that may make a practical difference to readers. Second, it is not clear how well the CWF will predict reading performances for students at various levels of reading proficiency. Researchers might examine reading performance in narrower or wider ranges of reading proficiency than that included in the present study.

Third, it is implied, but not demonstrated, that an intervention controlling the CWF of texts might increase the rate at which early readers develop reading proficiency. Although such a study is complicated to undertake, it is important to provide evidence on whether learning to read can be improved for some readers by systematic control of the CWF during instructional sequences. Fourth, further knowledge about the role of illustrations in texts for early readers would be useful. In particular, the manner in which illustrations engage young children in the task of reading when the reading task is challenging requires substantiation.

The CWF is based on the task demands that text features place on beginning readers. Understanding those demands, particularly for those whose literacy experiences occur primarily in school settings, is an essential research goal, and, ultimately, a source for informing policy in reading education.

NOTE

The research reported in this paper was supported, in part, by the U.S. Department of Education under Grant No. R203F990026.

REFERENCES

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

Adams, M. J., Bereiter, C., McKeough, A., Case, R., Roit, M., Hirschberg, J., et al. (2000). *Open court reading*. Columbus, OH: SRA/McGraw Hill.

Anderson, R. C., & Pearson, P. D. (1984). A schema-theoretic view of basic processes in reading comprehension. In P. D. Pearson, R. Barr, M. L. Kamil, & P. Mosenthal (Eds.), *Handbook of reading research* (pp. 255–292). New York: Longman.

Beck, I. L., & McCaslin, E. S. (1978). *An analysis of dimensions that affect the development of code-breaking ability in eight beginning reading programs* (Report No. 6). Pittsburgh, PA: Learning Research and Development Center.

Bruce, B. C. (1984). A new point of view on children's stories. In R. C. Anderson, J. Osborn, & R. J. Tierney (Eds.), *Learning to read in American schools: Basal readers and content texts* (pp. 153–174). Hillsdale, NJ: Erlbaum.

California English/Language Arts Committee. (1987). *English–language arts framework for California public schools (kindergarten through grade twelve)*. Sacramento: California Department of Education.

California English/Language Arts Committee. (1999). *English–language arts content standards 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.

Donahue, P. L., Finnegan, R. J., Lutkus, A. D., Allen, N. L., & Campbell, J. R. (2001). *The nation's report card for reading: Fourth grade*. Washington, DC: National Center for Education Statistics.

Farr, R. C., Strickland, D. S., Beck, I. L., Abrahamson, R. F., Ada, A. F., & Cullinan, B. E. et al. (2001). *Collections: Harcourt reading/language arts program*. Orlando, FL: Harcourt.

Flood, J., Hasbrouck, J. E., Hoffman, J. V., Lapp, D., Medearis, A. S., & Paris, S., et al. (2001). *McGraw-Hill reading*. New York: McGraw-Hill School Division.

Foorman, B. R., Francis, D. J., Davidson, K. C, Harm, M. W., & Griffin, J. (2004). Variability in text features in six grade 1 basal reading programs. *Scientific Studies of Reading, 8*, 167-197.

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

Gates, A. I., & Russell, D. H. (1938–1939). Types of materials, vocabulary burden, word analysis, and other factors in beginning reading. *The Elementary School Journal, 39*, 27–35, 119–128.

Gibson, E., & Levin, H. (1975). *The psychology of reading*. Cambridge, MA: MIT Press.

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.

Green, G., & Davison, A. (Eds.). (1988). *Linguistic complexity and text comprehension: Readability issues reconsidered*. Hillsdale, NJ: Erlbaum.

Hiebert, E. H. (2001). *An analysis of first-grade texts: Do the tasks differ across beginning reading programs?* (Tech. Rep. II-4.1). Honolulu, HI: NEARStar/Pacific Resources for Education and Learning.

Hiebert, E. H. (2005). State reform policies and the reading task for first graders. *Elementary School Journal, 105*, 245–266.

Hiebert, E. H., & Mesmer, H. (2005). Perspectives on the difficulty of beginning reading texts. In S. Neuman & D. Dickinson (Eds.), *Handbook of research on early literacy* (Vol. 2, pp. 935–967). New York: Guilford Press.

Hoffman, J. V., McCarthey, S. J., Abbott, J., Christian, C., Corman, L., & Dressman, M.. et al. (1994). So what's new in the "new" basals? A focus on first grade. *Journal of Reading Behavior*, 26, 47–73.

Hoffman, J., Roser, N., Patterson, E., Salas, R., & Pennington, J. (2001). Text leveling and little books in first-grade reading. *Journal of Literacy Research, 33*, 507–528.

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

Juel, C. (1988). Learning to read and write: A longitudinal study of fifty-four children from first through fourth grades. *Journal of Educational Psychology, 80*, 437–447.

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

Micro Power & Light. (1999). *Readability calculations* (Version 4.1). Dallas, TX: Author.

Minarik, E. (1957). *Little bear*. New York: Harper.

Modern Curriculum Press. (1997). *Ready readers*. Parsippany, NJ: Pearson Learning Group.

Morrow, L. M. (1986). Effects of structural guidance in story retelling on children's dictation of original stories. *Journal of Reading Behavior, 18*, 135–152.

National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel: Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading*. Washington, DC: U.S. Government Printing Office.

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.

Pinnell, G. S., Pikulski, J. J., Wixson, K. K., Campbell, J. R., Gough, P. B., & Beatty, A. S. (1995). *Listening to children read aloud: Data from NAEP's integrated reading performance record (IRPR) at grade 4*. Washington, DC and Princeton, NJ: Office of Educational Research and Improvement, U.S. Department of Education and Educational Testing Service.

Rhodes, L. K. (1979). Comprehension and predictability: An analysis of beginning reading materials. In J. C. Harste & R. F. Carey (Eds.), *New perspectives on comprehension* (pp. 100–131). Bloomington: Indiana University School of Education.

Rigby Education. (2000). *Rigby PM plus program*. Barrington, IL: Rigby Education.

Samuels, S. J. (1970). Effects of pictures on learning to read, comprehension, and attitudes. *Review of Educational Research*, 40, 397–407.

Shinn, M. R., Good, R. H., Knutson, N., Tilly, W. D., & Collins, V. (1992). Curriculum based measurement of oral reading fluency: A confirmatory analysis of its relation to reading. *School Psychology Review*, 21, 459–479.

Smith, D., Stenner, A. J., Horabin, I., & Smith, M. (1989). *The lexile scale in theory and practice: Final report*. Washington, DC: MetaMetrics. (ERIC Document Reproduction Service No. ED307 77)

Stein, M., Johnson, B., Boutry, S., & Borleson, C. (2000). *An analysis of the decoding instruction in six first-grade reading programs*. Tacoma, WA: University of Washington-Tacoma.

Stein, M. L., Johnson, B. J., & Gutlohn, L. (1999). Analyzing beginning reading programs: The relationship between decoding instruction and text. *Remedial and Special Education, 20*(5), 275–287.

Sulzby, E. (1985). Children's emergent reading of favorite storybooks: A developmental study. *Reading Research Quarterly, 20*, 458–481.

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.

Waterford Institute. (2000). *Waterford early reading program*. Scottsdale, AZ: Electronic Education.

Wright Group. (1996). *Sunshine reading program*. Bothell, WA: Wright Group/McGraw-Hill.