[3/15/02]

The Critical Word Factor in Texts for Beginning Readers: Effects on Reading Speed, Accuracy, and Comprehension[1]

Elfrieda H. Hiebert and Charles W. Fisher

Paper to be presented at Annual Meeting of AERA (New Orleans, April 4, 2002)

Running Head: Critical Word Factor & Beginning Readers

Abstract

This paper examines the effects of differences in the difficulty of texts on the speed, accuracy, and comprehension of beginning readers. Text difficulty was measured by the Critical Word Factor (CWF), an index of the word recognition demands of texts. The CWF is a function of the number of new, unique words per 100 running words of text that fall outside a designated group of high-frequency and phonetically decodable words. A common curriculum at the beginning of first grade consists of the 100 most-frequent words in written English and words that have a single grapheme representing a single phoneme (as in *go* and *cat*).

Thirty-six children completing their first trimester of first grade read four texts in a randomized order. Two texts had high CWFs, indicating a substantial portion of unique words beyond the 100 most-frequent words and words with relatively complex vowel patterns. Two other texts had low CWFs, indicating that unique words were limited to the designated curriculum of 100 most-frequent words or words with simple vowels.

Analyses of variance indicated that there were strong main effects for CWF on reading speed, accuracy and comprehension. All three variables were in the direction predicted by the model with the results for speed and accuracy being stronger than those for comprehension. Supplementary support for the model was provided by descriptive analyses. When means on speed, accuracy, and comprehension were examined by quartile (based on word recognition scores), 46 of 48 results were predicted by the model (two reversals

occurred on comprehension). In addition, words predicted by the model to be hard were hard and those predicted to be easy were easy.

The Critical Word Factor in Texts for Beginning Readers: Effects on Reading Speed, Accuracy, and Comprehension

The design of texts for instruction of beginning readers is a critical element in the development of reading programs. In the past several decades, the design of texts for beginning readers has undergone a series of major shifts. In 1987, the California English/Language Arts Committee opted for authentic literature for first graders as part of its textbook adoption guidelines and in 1990, the Texas Education Agency made a similar policy decision. In the succeeding round of textbook adoptions (California English/Language Arts Committee, 1999; Texas Education Agency, 1997), both states opted for decodable texts. These shifts from policies existing before 1987 to authentic literature in the 1990s to the current decodable texts represent dramatic swings in the features of texts used by beginning readers. Since textbooks developed for California and Texas are also those available to schools in other American states, these changes in text features were tantamount to a series of national interventions in a critical aspect of beginning reading instruction. Surprisingly, the research and theory that underlies these interventions is not extensive. While there is a voluminous literature on beginning reading, there are remarkably few studies that examine the effects of text features on students' reading speed, accuracy, and comprehension. Understanding how various features of text support (or hinder) children's reading acquisition appears to be an understudied aspect of research on beginning reading.

Much of the previous work on text features has combined selected features in an index or category system to allow comparison of one text with another. This typically results in assigning levels of difficulty to texts. This paper introduces a new measure of text difficulty called the critical word factor (CWF) based on the word recognition demands of texts for beginning readers. Data are presented on the effects of CWF on beginning readers' speed, accuracy, and comprehension in reading texts. This paper is the first in a series of studies exploring the validity of the CWF for indexing texts for beginning readers.

Existing frameworks of text difficulty

The dominant work on text difficulty for much of the 20th century centered on the notion of readability. Various text features like semantic complexity (e.g., presence of words on a designated list, or number of syllables per word) and syntactic complexity (e.g., number of words per sentence) were combined to create readability formulae. Generally speaking, readability formulae weighted the characteristics of texts much more than either the cognitive processes of the reader or the interaction of these two entities. Although the formulae provided useful results for relatively advanced readers, results for the earliest stages of reading acquisition were somewhat unstable. Take, for example, the popular text, *Little Bear* (Minark, 1957), that appears at the end of two of the five textbook programs that are currently approved in Texas. This text also appears on the approved lists of end-of-first-grade texts in many states. When readability formulae are applied, *Little Bear* yields readabilities that range from the middle of first grade to 7.5 (Micro Power & Light, 1999).

Critiques of readability formulas were not limited to instability at the early stages of reading. Beginning in the 1980s, the application of cognitive science perspectives showed that strict compliance to readability formulas could have negative consequences for comprehension (Bruce, 1984; Green & Davison, 1988). When high-frequency words were substituted for less frequent words or sentences shortened to lower readability indices, student comprehension also declined (Green & Davison, 1988).

These critiques of readability, among other factors, lead to a new generation of textbooks. For the 1988 state-wide textbook adoption, the California Committee for English/Language Arts (1987) mandated reading textbooks that consisted of authentic literature as opposed to texts that had been manipulated to comply with readability formulas. Three years later, the Texas Education Agency (1990) issued a similar mandate. Since cognitive scientists had not described text features that support children at the earliest stages of reading at the time these mandates were initiated, publishers were left with the task of identifying authentic literature that could be used with beginning first graders.

Hoffman et al.'s (1994) analysis of the texts adopted for use in Texas in 1993 showed a preponderance of the "predictable" text genre in first-grade textbooks. This 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 this point. Further, when the handful of existing studies was reviewed, they

showed that overuse of predictable texts failed to develop attention to word features, particularly among children at the early stages of word recognition (Hiebert & Martin, 2001).

As educators saw the effects of a diet of predictable texts during the beginning reading phase on subsequent reading strategies, alternative forms of texts were sought. The Texas Education Agency (1997) prescribed percentages of decodable words to be included in beginning reading textbooks. In their 1999 textbook guidelines, the California Committee on English/Language Arts implemented a similar prescription.

Scholarship on appropriate text features for beginning readers has not moved as rapidly as the policies of states and districts. However, four text difficulty schemes were identified in a recent review of literature (Hiebert, 2002): lexiles (Smith, Stenner, Horabin, & Smith, 1989), leveled texts (Fountas & Pinnell, 1999), predictability and decodability (Hoffman, Roser, Patterson, Salas, & Pennington, 2001), and potential for accuracy (Beck & McCaslin, 1978; Stein, Johnson, & Gutlohn, 1999). To illustrate these schemes, consider the difficulty ratings they generate when applied to *Little Bear*. This text is described by a lexile of 370, at the top of the 200-370 span that is recommended for grade one. 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.

Based on the criteria described by Hoffman et al. (2001), *Little Bear* would receive a predictability rating of 4 (somewhat predictable) and a decodability rating of 3 (both the

predictability and decodability scales have a range from 1-5). As expected, Hoffman et al. (2001) have found that better readers were able to read less predictable and less decodable texts. However, these ratings are not explicitly grounded in a framework describing the role of these variables at different points in children's reading development.

The potential for accuracy expands the scope of the earlier indices by considering not only text features but also what is expected to have been taught before a reader encounters a text. A student is considered to have potential for accuracy on a word if (i) the word is a high-frequency word that already has been taught or (ii) 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 is taken from the teacher's manual rather than from actual instruction (for practical reasons). By necessity, potential for accuracy was assessed in this way for all of the first-grade reading programs that were submitted for adoption in Texas where potential for accuracy levels were prescribed as 51% to 80% (Stein, Johnson, Boutry, & Borleson, 2000). For a given student then, it may not be clear whether or not the instruction actually occurred, when it occurred, or in what context it occurred. Further, a given student may require several exposures to content to apply it independently. *Little Bear* has been used in this section for illustration of difficulty indices. Since, the potential for accuracy assigned to *Little Bear* depends on the teacher's manual for the program in which *Little Bear* is embedded, no numerical index is offered.

Although each of these approaches for assessing text difficulty offers some benefit, none provides adequate information about what readers need to know in order to read a particular text (Hiebert, 2002). While researchers of beginning reading have struggled to develop more useful characterizations of text features, texts used in schools have undergone rapid, wholesale changes. These changes have been so large that the cognitive tasks required of beginning readers from one decade to the next could easily be seen as differences of kind rather than degree. Consider, for example, Scott Foresman's reading textbook program, that has the longest publishing record in the field. From 1962 to 2000, the number of unique or different words (per text) has gone from 18 to 187 for the first 10 texts in the program (Hiebert, 2001a). This ten-fold increase in the number of unique words to be read represents a sea change in developmental expectations for beginning first graders. It would be difficult to find an acceptable research base for such an ambitious rate of word introduction for beginning first graders – particularly first graders whose success depends on the quality of their school experiences. To the contrary, a substantial percentage of American fourth graders are not attaining national standards (Donahue, Finnegan, Lutkus, Allen & Campbell, 2001), a pattern that apparently begins in first grade (Juel, 1988). The texts that initiate at least a significant portion of American students to formal reading acquisition need to be based on theory and research on linguistic and cognitive learning processes.

The Critical Word Factor

The Critical Word Factor (CWF) describes the task demands for recognizing words in beginning texts. This index assesses two aspects of a text: (a) the match of linguistic content in the text with the phonetically regular and high-frequency words that are associated with particular stages of reading development and (b) the demands on cognitive processing as represented by the number of different words that cannot be figured out with a stage's target linguistic knowledge. Two sentences taken from texts at equivalent points in two different beginning reading programs illustrate differential task demands on beginning readers:

Example 1: I can hop, run, and dig.

Example 2: I found my old, orange tiger.

While each sentence has the same number of words, they differ in the linguistic knowledge demanded of beginning readers to be proficient. In the first sentence, knowledge of simple vowel patterns—where a single grapheme represents a single phoneme as in *go* or *cat*—is sufficient to recognize all of the words. To become a fluent reader of the thousands of words in written English, readers must generalize consistent relationships between letters and sounds (Adams, 1990; National Reading Panel, 2000). In most readability formulas (e.g., Chall & Dale, 1995), texts with many easily decodable words can be evaluated as difficult because words such as *hop* and *dig* occur infrequently in written English. Yet unless children quickly grasp the alphabetic relationship, their success in reading will be limited (National Reading Panel, 2000).

The word *I* in both of the examples illustrates a second category of linguistic information with which beginning readers require facility. Pronunciation of the word *I* is the same as the pronunciation of the name of the letter "i." However this can't be generalized to all vowels that also appear as words (*i.e.*, the word *A* and the letter "a"). Zeno et al. (1995) identify *I* as the 25th most-frequent word in written English. Among this group of 25 words, half have vowel patterns that are irregular. Yet these words need to be recognized quickly if children are to be successful, since this group of 25 accounts for one-third of the total words in texts (Carroll, Davies, & Richman, 1971). This rapid word recognition depends on a set for diversity in letter-sound relationships (Gibson & Levin, 1975).

While the 25 most-frequent words are essential in creating texts, emphasizing words based on their frequency alone raises issues of the relative balance between high frequency and highly decodable words. The inclusion of words such as *old* and *found* in the second example demonstrates a historically common direction in American reading instruction (Gates & Russell, 1938-39; Gray, Monroe, Artley, Arbuthnot, & Gray, 1956). These two words are among the 200 most-frequent words. While other fairly common words occur with similar vowel patterns (*told, cold* and *round, ground*), these vowel patterns fall into the complex vowel category. Further, by including the word *orange*, beginning readers must differentiate three different sounds associated with the grapheme *o* in one six-word sentence. Clearly, the cognitive demands of reading the two sentences are remarkably different.

Another set of words in the second example requires additional linguistic information: *orange tiger*. High-meaning or high-imagery words such as these are emphasized in some reading textbook programs. A body of research suggests that beginning readers remember words as a function of their imagery value (Laing & Hulme, 1999; Metasala, 1999). For example, highly decodable words that have an easily associated meaning are recognized more readily than highly decodable words that are less meaningful. However, in the early stages of beginning reading when children are grappling with both high-frequency and decodable words, their ability to focus on the high-meaning words, especially those with multiple syllables, is unknown.

Graphic illustrations in books may allow children to make one-to-one matches between objects and words. This relationship is included in the early levels of a widely used scheme for leveling books (Fountas & Pinnell, 1999). Using illustrations to support early readers may make the immediate task of naming objects easier. However, the effect on later recognition of these words (without illustrations), especially for children at the earliest stages of reading, appears to be negligible (Johnson, 2000).

Because of the increased processing demands placed on beginning readers, words that are not easily decodable or highly frequent are identified as critical in determining word recognition. The number of such words in an instructional program (i.e., over an extended series of texts) directly influences the cognitive processing demands on beginning readers. Comparing reading texts before and after the transition to authentic literature in Texas (Texas

Education Agency, 1987), Hoffman et al. (1994) reported that the number of unique words had increased while the number of total words had decreased. In a broader analysis of several reading programs, Hiebert (2001a) has shown that the number of unique words, relative to the number of total words, has stayed fairly constant, the recent mandate for more decodable words notwithstanding. Among other things, it remains unclear how many highly decodable words can be figured out in a single text on their first encounter.

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

Betts's (1946) criteria for independent, instructional, and frustration levels of oral reading proposed can be implemented using the CWF. Betts described the independent level as 99% of word recognition accuracy, the instructional level as 95%, and the frustration level

as below 95%. These levels are analogous to texts with CWFs of 1, 5, and greater than 5 respectively. The efficacy of these levels of text difficulty is supported by the finding of Fisher, Filby, Marliave, Cahen, Dishaw, Moore, and Berliner (1978) that reading materials producing low error rates (2 to 5%) were positively related to students' reading achievement.

Most current approaches to text difficulty (including lexiles and leveled books) have reported very little data (in some cases, no data) on children's reading performances as a function of one or more of the difficulty indices. Hoffman et al.'s (2001) validation of the predictability and decodability measure is perhaps the best example. However, the results may be weakened by methodological concerns (see Hiebert & Martin, 2001 for a review). Both the design of textbooks and beginning reading instruction could benefit from more and better indices of text difficulty, especially indices with empirical support.

The Focus of the Present Study

This study explores the validity of the CWF at the point where text features likely matter most—the early stages of reading acquisition. Examining first graders' strategies when taught with different 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, the end of the first trimester of first grade was chosen as the point to examine the effects of different text characteristics on children's reading.

In this initial study, the effects of CWF are studied with existing texts. The texts are drawn from "little book" programs that are widely used in American classrooms (Hiebert,

2001b). These little books are typically advertised as leveled according to the guided reading criteria of Fountas and Pinnell (1999). The potential pool of texts also included programs of "decodable texts." The selection is described more fully in the methods section.

The study examines the reading speed, accuracy, and comprehension of first graders at the end of the first trimester on texts differing in CWF. Thirty-six students read 4 texts (2 low-CWF and 2 high-CWF) in a repeated measures design. The primary research question was "Do students read low-CWF texts faster, more accurately and with higher comprehension compared to high-CWF texts?"

Method

Sample

The current study examines the reading performances of students who are progressing within the expected or typical range for first grade. The texts chosen for the study had been designated by their publishers for use near the end of the first trimester of grade-one. Data were collected during a two-week span at the end of the first trimester of grade one in the 2001-2002 school year.

Students in the sample were from two schools in a medium-sized, western city. Both schools had approximately 40% of students receiving free or reduced price lunch. Approximately 25% of the children in each school were English Language Learners. This percentage was near that of the state's average (Donahue et al., 2001). Children were selected from six first-grade classrooms, four in one school and two in the other.

Pilot testing had shown that children who were unable to recognize a handful of high-frequency words performed at the earliest of Sulzby's (1985) book-reading stages, either naming known letters in the text or pretending to read a story. Since such responses indicated little about the relative difficulty of the two types of text, a word-recognition measure was used as a screening measure. Children who returned parental permission letters were given the screening measure.

The screening measure began with a list of 10 high-frequency words, representing the most frequent 25 words (Carroll et al., 1971). Only children who read at least 5 of the first 10 words correctly were included in the sample. Children who attained this level continued with the task until they failed to read 6 consecutive words on 7 subsequent 10-word lists representing progressively less frequent words from the 26th through the 200th on the Carroll et al. list. Of the 36 students who read 5 of the first 10 words correctly, scores ranged from 5 through 79 with a mean of 37 and a median of 31. Five children had scores higher than 70 and four had scores lower than 10. Of the 36 students who were successful on the screening measure, 15 were boys and 21 were girls. This group included five students who were English Language Learners. The sample, then, represented a broad range of reading proficiency but less than the full range found in these schools.

Materials

The word recognition measure (described above) was also used as a covariate in some analyses. The primary materials for the study consisted of four texts representing two levels

of difficulty (low-CWF and high-CWF). All of the texts were selected from little book programs. Existing texts were used because of the face validity of such texts for educators. Descriptive analysis of the text features of little book programs was the basis for selecting the particular texts (Hiebert, 2001b). Their publishers indicated that the texts represented a single text level according to the guided reading levels (Fountas & Pinnell, 1999). According to the guided reading levels, the texts were at Levels C-D, levels that correspond to the end of the first trimester of grade one. The first text, labeled "High Text 1," came from the Rigby PM program published by Rigby/Elsevier. "High Text 2" came from the Sunshine program published by Wright Group/McGraw Hill. Both Low Text 1 and Low Text 2 came from the Ready Reader program published by Modern Curriculum Press.

The texts were selected using several criteria (see Hiebert (2001b) from an earlier analysis that included the same texts). The first had to do with the CWF. The curriculum for the first grade begins with the 25 most-frequent words and an introduction to vowel patterns in CV, VC and CVC words. The curriculum for the second half of the first trimester extends to the 100 most-frequent words and CCVC and CVCC patterns.

The texts used in Hiebert's (2001b) earlier analysis were reanalyzed to determine text characteristics for individual texts (the original analysis focused on sets of 10 texts). The four programs included Harcourt Collections (Farr et al., 2001), Waterford, Sunshine/Wright, and Rigby PM. When assessed against this curriculum of the 100 most-frequent words and vowel patterns in single-syllable words through CCVC and CVCC patterns, the average

CWF across 40 texts for this time period in grade 1 was 22 with a range from 6.6 to 47.9 and a median of 17. The texts from these programs with a CWF between the median (17) and the mean (22) constituted the pool from which the "high-CWF" texts were selected.

Texts with low-CWF contained no words beyond the curriculum designated for the first trimester of grade 1. The pool of texts from which the low-CWF texts were selected came from the Ready Readers program.

From these two pools, four texts were selected (2 high-CWF and 2 low-CWF). In making the final selections, two factors related directly to comprehension were also considered. Texts had to have sufficient content to ensure that a meaningful assessment of comprehension could be obtained. Since texts were to be presented to students without illustrations, the meaning of the text had to be apparent from the words alone. Using these criteria four texts were selected: one from the Wright Group (CWF equal to 21), one from Rigby's PM Readers (CWF equal to 20), and two from Ready Readers (CWF equal to 0).

Minor modifications were made to the selected texts. Several words were deleted from the texts in an attempt to keep the total number of words equivalent across the four texts. Each text contained approximately 50 words. The titles of the texts were also modified to make them similar for low- and high-CWF texts. Since titles were read to students as part of the experimental procedure, 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 provided in Table 1. The characteristics of the four texts are summarized in Table 2.

Insert Tables 1 and 2 about here

The texts were presented to students without illustrations. We wanted reading performances to depend on knowledge about text independent of clues that might be available from illustrations. However, a single illustration appeared on the cover of each text along with the title. Each illustration was chosen from a clipart library to be generically related to the text without divulging any of the specific content.

The primary differences between the two sets of books were the numbers of unique words, the average decodability of words, and the high-frequency ratings of words.

Procedures

The procedures used in collecting student responses to the word recognition measure and the four passages underwent two rounds of pilot testing. Each child participated in two short sessions over a two-day period. All sessions were one-on-one and conducted in a convenient area in the school but outside the child's classroom.

The first session began with the screening measure. If children did not read at least 5 of the first 10 words, they were thanked and returned to their classrooms. Children who did

get 5 of the first 10, continued until they missed 6 in a row or they reached the 80th word card. Responses for each word were recorded as correct or incorrect.

After the word recognition task was completed, children were given one of the four experimental texts. Each child read all four passages over the two sessions. During the first session a child read one low- and one high-CWF passage. In the second session, the child read the other two passages (one high- and one low-CWF) with the low-high ordering reversed. Since there were two examples of low- (and high-) CWF passages, reading order effects were still possible. To guard against such effects, the 8 possible orders (within the constraints described above) were listed and used sequentially during data collection, ensuring that the reading orders were balanced.

The data were collected by two researchers. One researcher worked with a child during the first session and the second researcher worked with that child during the second session. Half of the children read with researcher "a" for the first session and half 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.

Presentation of a text began with the researcher reading the title to the student. Students were then asked to continue reading the text. At this point, the researcher began timing how long it took the student to read the text. As the student read, the researcher

recorded the student's miscues, focusing on omissions, substitutions, and insertions. The researcher also wrote down the start and stop time of students' reading of each text. Following the completion of the student's reading, the student was asked the question, "Can you tell me what the story was about?" followed by the prompt "Can you tell me anything else that happened?" Student responses were written down verbatim. In the second session, the student met with the second researcher and read two additional texts. The same procedures of note-taking and tape-recording were followed.

Scoring. Students' readings of each of the four texts were assessed for speed and accuracy and their responses to the comprehension prompts were analyzed for level of comprehension. Since each child read 4 passages, there are 4 variables for each of the three measures of speed, accuracy, and comprehension for each child.

The speed variable is defined as number of words in the passage divided by the time elapsed for reading the passage. The metric of the speed variable is words per minute. A student's reading errors are not taken into consideration for this variable.

The accuracy variable indexes how error-free a child's reading of a passage was. The score on a given passage was obtained 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 in the passage then dividing the result by number of words in the passage over 100. Accuracy, then, is: $ACC = ((N-E) \times 100)/N$

(where N = total number of words in the passage and E = number of errors + number of insertions). The metric of the accuracy variable is words correct per hundred words.

Comprehension was rated from a student's responses to the question "Can you tell me what the story was about?" and the follow-up prompt asked immediately following his or her reading of the story. A 5-point rating scale was developed to score responses. This scoring scheme was modeled on those used for scoring comprehension on the National Assessment of Educational Progress (e.g., Donahue et al., 2001). 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. A score of 2 was awarded for evidence of a few concepts or actions from the text beyond the title, while a score of 3 indicated evidence of the theme of the text, including a representation of most of the actions or concepts of the text. A score of 4 was awarded to responses that indicated full comprehension of the text in which at least some details from the text were used to elaborate the theme.

Two trained raters rated each response independently. Subsequently, differences were examined and when agreement could be reached a rating was amended (never by more than 1 category). When agreement could not be reached the ratings were not changed. The average correlation between raters for the original ratings was 0.82 and for the amended ratings 0.92. For analysis, the comprehension scores of the two raters were averaged.

Results

The effects of texts with different CWF's on reading performance were assessed by analysis of variance. Analyses of reading speed, reading accuracy and comprehension are presented in turn. Each of the analyses applied the SPSS algorithm for repeated measures. For each outcome variable, three contrasts were examined: comparison of means for the two low-CWF texts; comparison of means for the two high-CWF texts; and comparison of the sum of the means for low-CWF texts and the sum of the means for the high-CWF texts. In each analysis, no differences were expected for the first two contrasts (i.e., no differences in performances on texts with the same CWF) but the third contrast was expected to yield differences (i.e., average performance on the low-CWF texts was expected to differ significantly from average performance on the high-CWF texts). Analyses of covariance, with word recognition as the covariate, were also run. The pattern of results was highly similar for both sets of analyses. Since the more complex analysis (ANCOVA) did not change the pattern of results, only the ANOVA's are reported here.

Following the analyses of variance, mean performances on the dependent variables are explored graphically. The sample was divided into quartiles on word recognition and then means on the dependent variables were calculated for the quartile groups. These means are presented as a series of bar graphs that are intended to complement the ANOVAs. Finally, errors on individual words were compared to the model's prediction of "hard" and "easy" words, ratings of decodability and an index of word frequency.

The Effect of CWF on Reading Speed

Means and standard deviations for reading speed on each of the four texts are presented in Table 3. The ANOVA results indicate a strong main effect of CWF on reading

---

Insert Table 3 about here

---

speed: F (3, 105) = 44.0 (p < .001). The results for the three contrasts were as expected. Contrasts for pairs of texts with the same CWF (i.e., low1 to low2) were not significant (>.10), indicating no differences in reading speed. The contrast between the two pairs (i.e., lows to highs) was significant (p < .01).

The Effect of CWF on Reading Accuracy

Means and standard deviations for reading accuracy on each of the four texts are included in Table 3. The ANOVA results indicate a strong main effect of CWF on reading accuracy, F (3,1 05) = 35.5 (p < .001). The first (the 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 was significantly different for the two high-CWF texts (p < .01). Despite this unexpected difference between the two high-CWF texts, the means for accuracy on both high-CWF texts were lower than those for low-CWF texts as expected. The mean for reading accuracy was

the same for the two low-CWF texts: 86.7%, while the mean for the one high-CWF text was 78.3% and 70.7% for the other high-CWF text.

The Effect of CWF on Reading Comprehension

Means and standard deviations for reading comprehension on each of the four texts are presented in Table 3. The ANOVA results indicate a significant main effect of CWF on reading comprehension: F (3, 105) = 10.9 ($p < .001$). The contrasts reveal that comprehension on like-CWF texts was approaching significance at the .05 level. However, the third contrast of differences across the texts by common type 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-CWF text 2 (See Me) where the mean was 2.2 and high-CWF text 2 (Up and Down) where the mean was 1.4 with the other two texts making no contribution. The low-CWF text 1 (Hop, hop, hop) and high-CWF text1 (My book) had the same mean of 1.8.

Mean Reading Performance by Quartile on Word Recognition

The 36 students in the sample were divided into quartiles (9 students per group) according to their scores on the word recognition measure. Figure 1 displays the average

---

Insert Figure 1 about here

---

word recognition score for each quartile. The means rise almost linearly from 10.0 for Q1 to 70.3 for Q4. Figures 2, 3, and 4 display quartile means for speed, accuracy, and comprehension (respectively) on each of the four texts.

Insert Figures 2 through 4 about here

For speed and accuracy (Figures 2 and 3), the means increase in more or less orderly fashion over quartiles with the low-CWF texts always higher than the high-CWF texts. These two variables follow the predictions of the model with the data on speed being somewhat stronger than for accuracy. There is also regularity in the means for comprehension in that the means for each text increase from Q1 to Q4 (see Figure 4). However, there are two reversals that are not predicted by the model. For Q2 and Q4, the mean for high-CWF-1 is greater than the mean for low-CWF-1. With these exceptions, results for speed, accuracy and comprehension are predicted by the model.

Reading Errors and Characteristics of Words

By examining each child's reading errors on the texts, additional information about the model can be obtained. For each word in each text, the model predicts whether words are easy or hard. According to the model, words that are beyond the high-frequency list associated with the curriculum and are not decodable are designated as hard words for these readers. In the four texts in this study, there were 20 hard words. At least one third of the

students made errors on 19 of these "predicted" hard words. The only exception was the word *book*. Although the word *book* is beyond the high frequency list for these students and is not decodable, it is not surprising that few children made errors in reading this word. Ninety-five percent of the words predicted to be hard by the model were hard.

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

The number of errors on each unique word in the texts was recorded. Each of the unique words also was indexed 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 0.64. Unique words were also indexed by frequency of occurrence. The correlation between errors per word and frequency of occurrence was 0.55.

Summary of Results

Analyses of variance indicated that there were strong main effects for CWF on reading speed, accuracy and comprehension. All three variables were in the direction predicted by the model with the results for speed and accuracy being stronger than those for comprehension. Supplementary support for the model was provided by descriptive analyses.

When means on speed, accuracy, and comprehension were examined by quartile (based on word recognition scores), with the exception of two reversals on comprehension, all results were predicted by the model. In addition, words predicted by the model to be hard were hard and those predicted to be easy were easy.

Discussion

This study represents the initial step in establishing the usefulness of an index of critical or difficult words in describing the tasks that texts pose for beginning readers. Data were gathered in 6 classrooms at the end of the first trimester in grade 1. Using word recognition on a high frequency word list as a screening device, 36 children were selected for the study. The sample constitutes a very broad band, but by no means the full range, of word recognition among students in mid-November of grade 1. For the children in the sample, the CWF predicted which texts would be difficult and which would be easy. The CWF of a text is a function of the high-frequency words and the decodability of words targeted in the curriculum. Children's performances on reading speed, accuracy, and comprehension of low- and high-CWF texts differed in the direction predicted by the model.

Beyond this basic finding, there are several points that are suggested, either directly or indirectly, by the study. In this study, the decodability of a given word predicted the number of reading errors quite well on that word. This occurred even though the children were learning to read in classrooms where explicit phonics was not a dominant instructional theme. Children appear 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.

The findings on reading speed may be particularly important in the design of texts for beginning readers. Analyses of sequential texts from widely-adopted, first-grade programs indicate that, whatever their philosophy, substantial numbers of new words are introduced. Hiebert, Martin, and Menon (in press) analyzed the texts in the literature and decodable components of Harcourt's Collections (Farr et al., 2001) and Open Court (Adams et al., 2000), and Reading Mastery (Englemann & Bruner, 1995), three prominent programs that are advertised as philosophically different from one another. Percentages of new, unique words per 100 were not substantially different across the first level of the same component of the three programs: The average number of new, unique words per 100 words was 31 for first level of the literature components and 27 for the first level of the decodable components. Almost 60% of the unique words in the literature components and 40% of those in the decodable components appear a single time over all of the texts for the first level. Further, the overlap between the words in the literature and decodable components within the first level of the same program was low with only 7 to 17% of a word corpus shared across two components. These results mean that, for every beginning reading text of 100 words, there

are approximately 30 new words, half of which will not be encountered in that instructional level again.

In subsequent levels of these components, except for the decodable books of Reading Mastery, the unique word, singleton, and shared vocabulary statistics remain the same. In subsequent levels, however, the percentage of multisylllabic words and monosyllabic words with complex vowels increase substantially. The time students require for figuring out words in these texts is likely to increase. When figuring out new words takes an inordinate amount of time, it is little wonder that a substantial number of children within an age cohort fail to develop sufficient speed in reading text (Pinnell, Pikulski, Wixson, Campbell, Gough, & Beatty, 1995).

However, it is not only reading speed that is likely to stagnate or decline. As students work longer and harder to figure out more words, comprehension is also likely to deteriorate (Shinn, Good, Knutson, Tilly, & Collins, 1992). Learning to read is likely to continue to be a remarkably complex phenomenon. Maintaining fluency at an appropriate level, especially at the early stages of learning to read, may be particularly important. Without attention to fluency considerations in the design and selection of texts, too many students will rarely see texts that they can read with speed and comprehension.

This study provides evidence that the CWF of a text can be used to predict important student reading outcomes (reading speed, accuracy and comprehension). As a result, CWF, focusing more closely on the demands of word recognition, may be a more useful index of

difficulty for the early stages of reading than indices with a considerably broader focus. Consider, for example, the widely used guided reading levels (Founts & Pinnell, 1999). This popular index includes a variety of criteria such as naturalness of language and overall length of text. All four texts used in this study are designated "C-D" by the guided reading procedure. However, children performed very differently on the low- versus high-CWF texts. Hence, the broader definition of the guided reading levels may erode their usefulness for students similar to those included in this study.

On the other hand, illustrations are a key component of guided reading levels but the current study presented texts without illustrations. In removing the effects of illustrations, we may have increased the apparent differences between CWF and guided reading levels. However, a heavy reliance on illustration, especially when the number of difficult words is high, is unlikely to help children attend to word-level features (Samuels, 1970). There is no attempt to downplay the role of illustrations here. Rather, we want to foster understanding of a variety of factors that influence children's reading performances.

Although it is very early days in the exploration of CWF as an index of text difficulty, the initial results appear strong enough to encourage additional research. Four avenues are suggested for subsequent studies. First, it is not clear what increment in CWF will yield practical results for readers and texts. The current study compared texts that differed in CWF by a relatively large amount. Subsequent studies could examine smaller differences in CWF to explore the minimum difference in CWF that can be expected to

make a practical difference to readers. Second, it is not clear how well CWF will predict reading performances for students at various levels of reading proficiency. New studies may be designed to 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 skill. While such a study is much more complicated to undertake, it is important to provide evidence on whether or not learning to read can be improved by systematic control of CWF during instructional sequences. Fourth, it would be useful to know more about the role of illustrations in texts for early readers. In particular, the manner in which illustrations engage young children's in the task of reading, at a point where the reading task is challenging, requires substantiation. At the same time, the forms of illustrations that do not impede independent word recognition also require documentation.

To date, our effort to describe the effects of text features on beginning readers' achievement has concentrated on word recognition demands. We recognize that this model requires further elaboration to account for factors that engage readers and that influence interpretations of text. We believe, however, that the attention that this work directs to the tasks that texts pose for beginning readers, particularly those whose literacy experiences occur primarily in school settings, is essential as part of research and, eventually, policy in reading education.

Footnote

[1]Supported in part by the U.S. Department of Education under Grant Number R203F990026 administered by Pacific Resources for Education and Learning.

References

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

Adams, M.J., Bereiter, C., Brown, A., Campione, J., Carruthers, I., Case, R., Hirschberg, J., McKeough, A., Pressley, M., Roit, M., Scardamalia, M., & Treadway, G.H., Jr. (2000). Open Court Reading. Columbus, OH: SRA/McGraw-Hill.

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

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. Berliner (1981)

Betts, E. (1946). Foundations of reading instruction. New York: American Book.

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 reading in American schools: Basal readers and content texts (pp. 153-174). Hillsdale, NJ: Erlbaum.

California English/Language Arts Committee (1999). English-Language Arts Content Standards for California Public Schools (Kindergarten Through Grade Twelve). Sacramento: California Department of Education.

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

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

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

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.

Engelmann, S., & Bruner, E.C. (1995). Reading Mastery. Columbus, OH: SRA/McGraw-Hill.

Farr, R. C., Strickland, D. S., Beck, I. L., Abrahamson, R. F., Ada, A. F., Cullinan, B. E., McKeown, M., Roser, N., Smith, P., Wallis, J., Yokota, J., & Yopp, H. K. (2001). Collections: A Harcourt Reading/Language Arts program. Orlando, FL: Harcourt School Publishing.

Fisher, C.W., Filby, N.N., Marliave, R., Cahen, L.S., Dishaw, M.M., Moore, J.E., & Berliner, D.C. (1978). Beginning teacher evaluation study (Technical Report, V-1). San Francisco, CA: Far West Laboratory.

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. (1938-39). 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. (2002). Standards, assessment, and text difficulty. In A. E. Farstrup & S. J. Samuels (Eds.), What research has to say about reading instruction (3rd Ed., pp. 337-369). Newark, DE: International Reading Association.

Hiebert, E.H. (2001a, April). Pace and repetition: The forgotten variables in the design of beginning reading programs. Paper presented as part of the symposium Texts that support beginning reading acquisition: What we know and how we know it at the annual meeting of the American Educational Research Association, Seattle, WA.

Hiebert, E.H. (2001b). An Analysis of First-Grade Texts: Do the Tasks Differ Across Beginning Reading Programs? (Deliverable II-4.1). Honolulu, HI: NEARStar & Pacific Resources for Education and Learning.

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

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

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

Hoffman, J., 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.

Klare, G. (1984). Readability. In P. D. Pearson, R. Barr, M. Kamil, & P. Mosenthal (Eds.), Handbook of reading research (pp. 681–744). New York: Longman.

Laing, E., & Hulme, C. (1999). Phonological and semantic processes influence beginning readers' ability to learn to read words. Journal of Experimental child Psychology, 73, 183-207.

Metsala, J.L. (1999). Young children's phonological awareness and nonword repetition as a function of vocabulary development Journal of Educational Psychology, 91, 3-19.

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

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

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

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: Office of Educational Research and Improvement, U.S. Department of Education; Princeton, NJ: 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, IN: Indiana University School of 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 number ED 307 577).

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.

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.

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.

Zeno, S.M., Ivens, S.H., Millard, R.T., Duvvuri, R. (1995). The educator's word frequency guide. New York: Touchstone Applied Science Associates, Inc.

Table 1. Examples of low- and high-CWF texts

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

Table 2. Characteristics of Texts

	Total Words	Unique words (Unique words per 100)	High-Frequency Rating	Decodability Rating	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

Table 3. Descriptive statistics for reading speed by text, reading accuracy by text, and reading comprehension by text

Text	N	Speed		Accuracy		Comprehension	
		Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Low 1 (Hop Hop Hop)	36	47.0	24.7	86.7	14.3	1.8	0.9
Low 2 (See Me)	36	48.5	26.8	86.7	14.2	2.2	0.8
High 1 (My Book)	36	30.9	24.8	78.3	18.3	1.8	0.8
High 2 (Up and Down)	36	28.1	17.3	70.7	18.9	1.4	0.7

Figure 1. Means for word recognition by quartile within the sample.

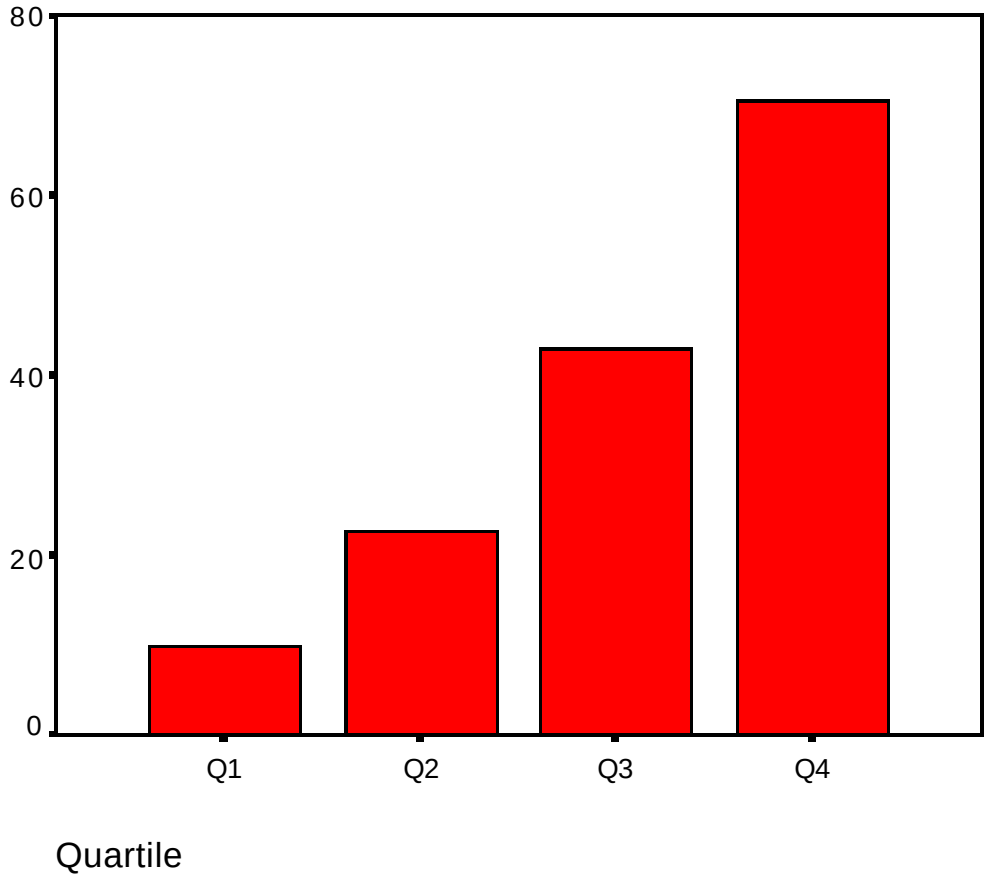

Figure 2. Means for reading speed by quartile on four texts.

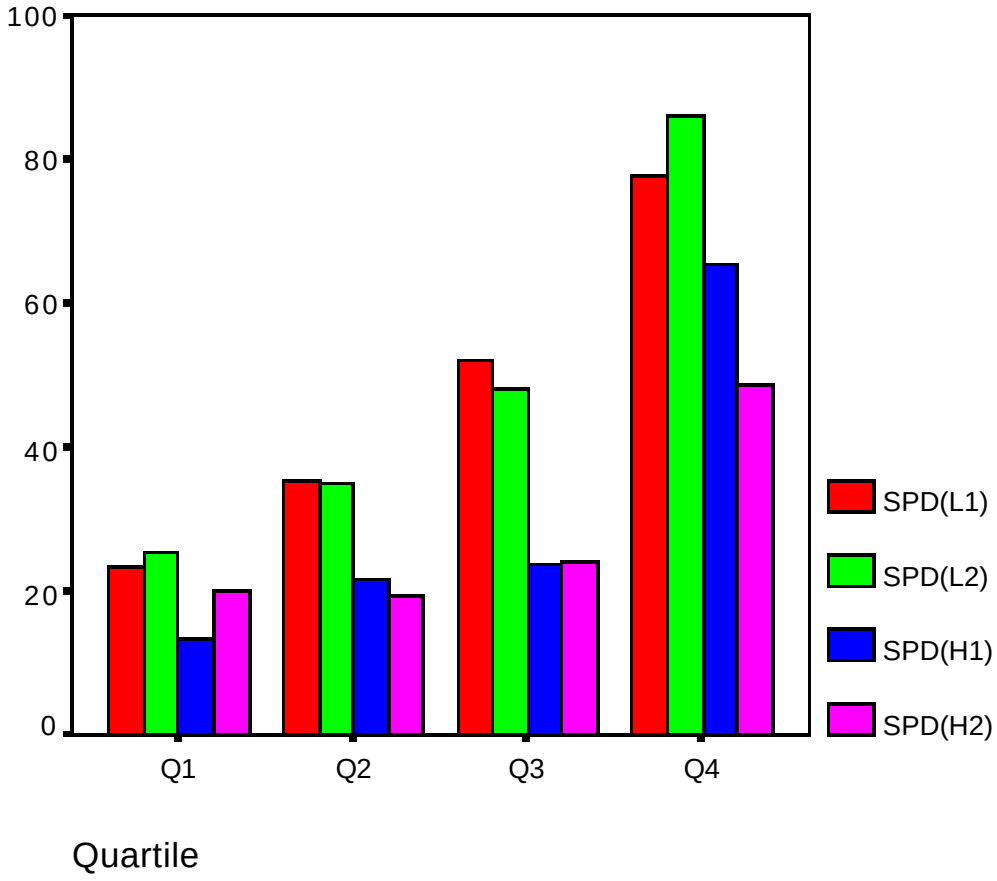

Figure 3. Means for reading accuracy by quartile on four texts.

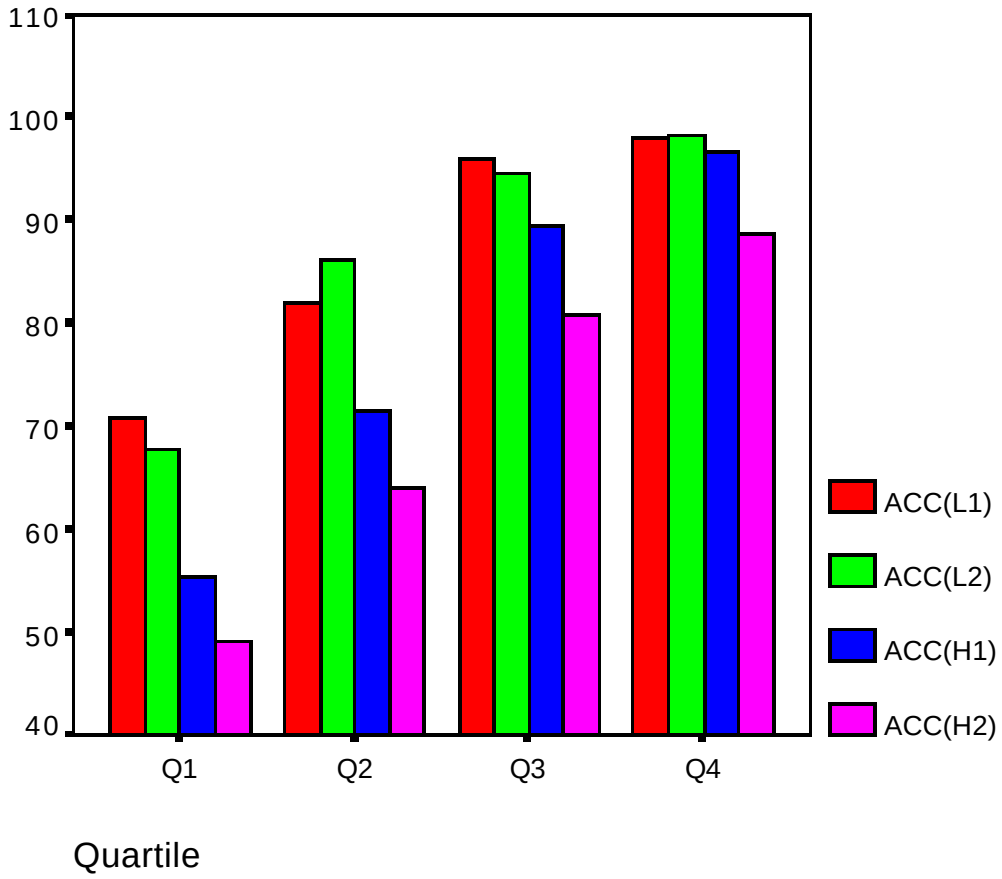

Figure 4. Means for reading comprehension by quartile on four texts