# 27

# *Perspectives on the Difficulty of Beginning Reading Texts*

❖

*ELFRIEDA H. HIEBERT*
*HEIDI ANNE E. MESMER*

Features of the text influence the quality of the interaction that can occur between a reader and a text. For beginning readers who know only a handful of words (and likely idiosyncratic ones such as their names), most texts will require scaffolding by a proficient reader for an interaction with text to occur. Choosing texts for instruction that support beginning readers in moving from scaffolded to independent reading has been one of the most persistent challenges facing teachers and teacher educators.

The critical role of success in beginning reading has meant that considerable attention has focused on selecting texts. Over the past two decades, agencies in numerous American states have taken on the task of selecting appropriate texts for beginning readers in their jurisdiction. Of the four largest American states, three (California, Texas, and Florida) identify basal reading programs that are acceptable for use with state funds. The inclusion of programs on these state lists is often predicated on compliance with guidelines on the features of beginning-level texts. The systems for sorting first-grade texts would seem to be a first point for examining the scientific foundation that is the byword of current federal policies. The practices are critical, and the investments in state and federal dollars and of teacher and student time are high.

In this chapter, we review theory and research on schemes that are used currently to determine the difficulty of texts at the beginning levels of reading. In doing so, this chapter builds on and extends the work of Hiebert and Martin (2001), which reviewed the features of words and how they are learned. The three primary text-difficulty methods—readability, guided reading levels, and task-based systems—apply particular assumptions about what is critical in beginning reading acquisition. As a result, each yields a different index on beginning reading texts. To illustrate the data that the text difficulty systems produce, we have chosen a prototypical text that exemplifies a second-trimester, first-grade text from a program based on the text-difficulty system. For one of the text difficulty schemes—readability formulas—we have selected two prototypical texts: one for conventional readability formulas (e.g., Spache, 1981) and the other for lexiles, a current manifestation of readability formulas (Smith, Stenner, Horabin, & Smith, 1989). Excerpts from these four texts and a fifth, *The Cat in the Hat* (Geisel, 1957)—described by Anderson, Hiebert, Scott, and Wilkinson (1985) as an ideal first-grade text—appear in Table 27.1.

Following the scheme's ratings of the five texts, three aspects of a text-difficulty method are described: (1) its rationale and a

**TABLE 27.1. Ratings According to Six Text-Difficulty Systems of Five First-Grade Texts**

| Text difficulty scheme | Excerpt from a prototypical text | Readability | | Text leveling | | Task-based | | Total words |
|---|---|---|---|---|---|---|---|---|
| | | Spache | Lexile | Guided reading | STAS-1 (predictability, decodability ratings) | Decodability | CWF | |
| Readability | Dad looked at Molly's red nose. "You will have to go to bed," said Dad. "You have a cold." "I don't have a cold!" said Molly, blowing her nose. But she went up to bed and went to sleep with her big red cat at her side. "I put Molly to bed," said Dad. (Cummings, 1983) | 1.8 | 310 | H | 3.5 (4, 3) | 1.7 | 2.6 | 426 |
| Lexile | "Do you have a bed just right for a pig?" he asked the saleslady. "Hmmmm," she said, looking Poppleton over. "Right this way." Poppleton followed the saleslady to the biggest bed in the store. It was vast. It was enormous. "It's just my size," said Poppleton. (Rylant, 1998) | 2.0 | 250 | J | 4.5 (5, 4) | 2.0 | 8.4 | 665 |
| Guided reading levels | She flew the hang-glider at the school picnic. We drew pictures of her.<br>The next day, Mrs. Bold came to school with a broken arm. Look at Mrs. Bold! How did you break your arm? Were you driving the rally car? What happened? Did your hang-glider crash? (Beck, 1993) | 2.2 | 290 | F | 3.5 (3, 4) | 1.9 | 18 | 93 |
| Task-based | "You must have a fever and a cold." "Dragons don't get colds," creaked Dee. "Dragons breathe hot flames." "Can you breathe flames?" asked Dad. "No," creaked Dee. Dad made a pot of tea. "This tea's heat will help you breathe," he said. "Dragons don't like tea," creaked Dee. (Raymer, 1993) | 2.9 | 270 | G | 3.5 (4, 3) | 1.6 | .8 | 127 |
| Prototype primer text (Anderson et al., 1985) | Then he got up on top with a tip of his hat. "I call this game FUN-IN-A-BOX," said the cat. "In this box are two things I will show to you now. You will like these two things," said the cat with a bow. "I will pick up the hook." (Geisel, 1957) | 2.4 | 270 | J | 3 (4, 2) | 1.6 | 1.4 | 1,625 |

brief history, (2) a review of empirical investigations on the reliability and validity of the system, and (3) conclusions about its strengths and weaknesses. The principle that drives the latter discussion is the usability of the information that a system supplies for teachers' use in knowing what to teach their students. We begin with the system that has the longest history—readability formulas—and then move to the two systems that have replaced readability formulas in many published programs—guided reading levels and systems that are based on tasks such as decodability.

## Readability Formulas

The prototypical text from the era when readability formulas were used to vet texts, *Molly's Surprise* (Cummings, 1983), had a readability of 1.8 ( Spache, 1981), a level very close to the 1.75 associated with the second trimester of grade 1. The other four texts had readabilities of between 2.0 and 2.9. From the readability perspective, *Dragons Don't Get Colds* (Raymer, 1993) was one grade level more difficult than *Molly's Surprise*.

The text at the 250 lexile level that corresponds with the second trimester of grade 1 (*Scholastic Reading Inventory*, 2002) is *Poppleton Everyday* (Rylant, 1998). In that grade levels are typically evaluated in terms of units of 200 on the lexile scale (Smith et al., 1989), the lexiles of the other four texts are within a narrow and comparable range to the prototypical second-trimester text of 250 lexile: 270–310.

### *Description of and Rationale for Readability Formulas*

Although Lively and Pressey's (1923) proposal for the measurement of vocabulary burden in school textbooks is typically identified as the first readability formula, it was Gray and Leary's (1935) formula that provided the paradigm for readability formulas for the subsequent 60 years. From 289 factors that 100 experts and 100 library patrons identified as possible contributors to readability, Gray and Leary selected 44 that could be counted reliably and that occurred with sufficient frequency in their criterion passages (the Adult Reading Test). Using the scores of poor adult readers on these passages, Gray and Leary identified five variables that accounted for a sufficient amount of variance on a multiple regression analysis ($R = .65$): (1) number of "hard words" not on a list of 769 words; (2) number of personal pronouns; (3) average number of words per sentence; (4) percentage of different words; and (5) number of prepositional phrases. Over the next five decades, researchers reduced the number of variables to two or three, but Gray and Leary's basic procedure became the model for numerous readability formulas.

As Klare noted in his review in 1984, formulas vary in the data that they provide on the same text as a function of developmental criteria and the range of readers' ability on the criterion task. The formula that has been described as most valid for primary-level texts, both currently (Good & Kaminiski, 2002) and historically (Klare, 1984), is the one developed by Spache (1953, 1981). Spache's formula used two dimensions of texts: sentence length and the percentage of total words that were difficult words (i.e., not on a list of 1,040 words that Spache identified from analyses of textbooks). The underlying perspective on text difficulty can be illustrated by slight alterations to the primer-level text, *Molly's Surprise* (Cummings, 1983): (1) breaking several sentences into shorter ones and (2) changing Molly's name to Penny. When *Molly's Surprise* becomes *Penny's Surprise,* the readability of the text changes from 1.8 to 1.5. Passages with fewer words per sentence are deemed easier for beginning readers than those with longer sentences. *Penny* is on Spache's list of 1,040 words that primary-level students are to know; *Molly* is not. According to Davison and Kantor (1982), readability formulas were used to create school texts, not simply to adjust texts to comply with formulas. For example, because words such as *ice cream, picket fence, milkman*, and *castle* were on the Spache or a similar list, writers for textbook programs developed stories with these words.

As a result of research from a cognitive science perspective in the 1980s (see, e.g., Davison & Kantor, 1982) that identified

problems with texts that had been manipulated or written to satisfy readability constraints, the field's two primary professional associations, the International Reading Association and the National Council of Teachers of English, called for cautious use of readability formulas (Michelson, 1985). This call was echoed in *Becoming a Nation of Readers* (Anderson et al., 1985), in which a moratorium on the use of readability formulas was advocated. Such initiatives led to a decrease in the use of readability formulas and an increase in alternative text-difficulty schemes such as guided reading levels and decodability.

Although widely used textbook programs still appear to use readability formulas sparingly, if at all (see, e.g., Pikulski, 2002), two activities are drawing attention back to readability formulas. One is the use of readability formulas within prominent assessments such as the Dynamic Indicators of Basic Early Literacy Skills (DIBELS; Good & Kaminski, 2002) which was reported to be used in over 1 million kindergarten through third-grade classes during the 2003–2004 school year (see the DIBELS website, http://dibels.uoregon.edu/). According to Good and Kaminski, the difficulty of texts on the DIBELS was validated by Spache's readability formula.

A second source for renewed interest in readability formulas is the presence of several computer-based readability programs, such as lexiles (Smith et al., 1989) and ATOS (School Renaissance Institute, 2000). The lexile framework, for example, orders texts according to a scale from 0 to 2000, with beginning texts at the lower end and graduate school and technical texts at the upper end. Although the reporting units of lexiles or ATOS are different from the grade levels of conventional readability formulae, their criteria are syntactic complexity, as measured by sentence length, and semantic complexity, as measured by the number of words that fall within anticipated bands of words (Smith et al., 1989; School Renaissance Institute, 2000).

The digital technology that underlies this new generation of readability systems makes it possible to base text levels on large corpora of words. Further, words that have become archaic, such as *milkman*, are relegated to rare-word status. At the same time, educators who use these digital readability systems do not have access to the corpora that are associated with particular text levels, as was the case with the Spache. Whereas the words that designate *Poppleton Everyday* (Rylant, 1998) as a second-grade text according to the Spache can be identified (e.g., *enormous, saleslady, crackers, bluebirds, pillows),* teachers, students, and parents are given no guidelines as to the vocabulary that underlies the designated lexile. Presumably the words that account for the lexile rating and the Spache level are similar. However, without data from the readability developers, this conclusion can only be inferred.

### *Empirical Evidence for Readability Formulas and Beginning Readers*

According to Chall (1988), readability formulas at the primary level originated with studies of vocabulary control in the 1920s and 1930s that examined the number of new words per book, their repetitions, and their frequency. Researchers assumed that texts with high numbers of new words with few repetitions and/or low frequencies in written English created obstacles for reading acquisition. However, as Chall (1988) has observed, only one experiment had been conducted before these assumptions were used to create textbooks. This single study—that of Gates (1930)—considered the optimal number of repetitions for first graders of different ability levels. Based on a rather limited sample of text, Gates concluded that average-ability students (as defined by IQ) required 35 repetitions of high-frequency words and, from these data, extrapolated the number of repetitions required by high-achieving and low-achieving students.

Whereas experimental studies were infrequent, studies pertaining to the validity of different readability formulas were numerous. In 1984, Klare stated that over 1,000 studies had been conducted on readability. Many of these studies examined the variables that accounted for readers' performances on a set of passages, often the McCall–Crabbs Standard Test Lessons in Reading (McCall & Crabbs Schroeder, 1926/1979). Numerous other studies reported on the concurrent validity of a new set of variables relative to existing formulas. Syllable counts were proven to be valid by

showing strong correlations with earlier formulas that used other measures of semantic complexity (Klare, 1984). Or lexiles were described as valid on the basis of strong correlations between lexile levels and graded texts within textbook programs that complied with conventional readability formulas (Smith et al., 1989).

The circuitous process whereby formulas were based on a set of passages that had been developed according to the same criteria as the formulas meant that texts could be ordered across a set of grades with consistency. However, the narrower the band of performance, the more difficult it was to make differentiations. Even more challenging was the task of applying the criteria of semantic complexity at the very earliest grades. The 10 readability formulas that are part of the Micro Power and Light (1999) software produced readabilities for *Molly's Surprise* that ranged from .6 to 5.7 grade levels and, for *Penny's Surprise*, from 0 to 4.6. The only formula to make a fine-tuned distinction across these two passages was the Spache (1981). In that the original and revised texts had been written to comply with the Spache formula, this finding should not be surprising.

In the 1980s, perspectives from cognitive science and linguistics were applied to the texts that resulted from this circular process of developing and validating readability formulas. These analyses showed that, when high-frequency words were substituted for less frequent but more descriptive words, meanings of texts were changed and even made more obscure (Davison & Kantor, 1982). By shortening sentences to comply with readability formulas, conjunctions were often eliminated, and causal connections between ideas were obscured. When comprehension of unmanipulated texts was compared with that of manipulated texts, students' superior performance on the former was taken as evidence that readability formulas were detrimental to effective comprehension (e.g., Beck, McKeown, Omanson, & Pople, 1984). In none of these studies of restructured texts, however, was the focus on beginning readers. No studies compared, for example, beginning readers' proficiency with texts with no vocabulary restrictions relative to texts with a modicum of vocabulary control.

### *Conclusions: Readability Systems*

Even with extensive computerized databases (Smith et al., 1989), readability systems continue to be limited in their support of instruction. A grade level of 1.8 on the Spache (1981) does not indicate the proficiencies with which readers must be facile to read this level text. Neither does a lexile level of 250 indicate what beginning readers need to know to move to a higher level. However, when the new readability systems are evaluated relative to the old readability systems, the latter are more useful for instruction and assessment than the former. At least with a formula such as Spache's (1981), the words that are associated with the primary grades are known. This information is valuable for educators as they respond to policy mandates that are based on tests that use the Spache. The five texts in Table 27.1 differ in the distribution of high-frequency words and multisyllabic words, among other features. However, the lexile system does not distinguish between these texts in any discernible way, and little information is forthcoming from the system as to the underlying curriculum.

## Text Leveling Systems

From the vantage point of current U.S. classrooms, the most widely used text-difficulty scheme consists of Fountas and Pinnell's (1996, 1999, 2001) guided reading levels. A text at the third of the four levels associated exclusively with grade 1—level F—was selected from available lists: *Mrs. Bold* (Beck, 1993). Fountas and Pinnell (1996, 1999) also provide levels for *The Cat in the Hat* and *Poppleton Everyday.* A reading specialist with a decade of experience in text leveling established the levels for the two remaining texts. As Table 27.1 shows, three were within the exclusive first-grade range, and two had second-grade levels. *The Cat in the Hat* was among the latter, with a level of mid- to late second grade.

Hoffman, Roser, Patterson, Salas, and Pennington (2001) have also developed a text-leveling system, the Scale for Text Accessibility and Support (STAS-1). Ratings of the texts according to the STAS-1 are included in Table 27.1. These ratings indicate

that differentiation across the texts is not substantial with this 5-point scale. Three of the texts had the same rating (although they had somewhat different distributions, according to predictability and decodability scales).

### *Description of and Rationale for Text-Leveling Systems*

The leveling of texts by experts or judges is not a recent phenomenon (see, e.g., Carver, 1976; Singer, 1975). However, this procedure was not prominent until readability formulas were eliminated as a criterion for textbook selection in America's largest states (California English/Language Arts Committee, 1987; Texas Education Agency, 1990). The theory underlying the use of literature and little books in textbook programs posited that readers employ multiple sources of information in understanding unknown words, including the structures of syntax and texts (Goodman, 1968). The Reading Recovery levels that have evolved into the guided reading levels were a response to this need. Like the guided reading levels, the STAS-1 (Hoffman et al., 1994) uses experts' judgments or ratings. Unlike guided reading levels, which are presented as a holistic score, the STAS-1 gives ratings on individual categories. Consequently, the uniquenesses of the two schemes will be developed.

#### READING RECOVERY/ GUIDED READING LEVELS

As little books became prominent in school reading programs, particularly as Reading Recovery programs were initiated in U.S. schools during the mid- to late 1980s, Peterson (1991) developed a scheme for establishing text difficulty of the little books. Similar to the primary trait model of holistic scoring that has a long history in writing assessment (Cooper & Odell, 1977), four dimensions were identified as the basis for a text level: (1) book and print features; (2) content, themes, and ideas; (3) text structure; and (4) language and literary elements. Unlike primary trait schemes in writing, however, the four separate dimensions of the guided reading levels were not analyzed individually. A single score was provided with no indication of the weight or scoring of individual dimensions.

Within guided reading levels, Fountas and Pinnell (2001) have extended the original four criteria (book and print features, content, text structure, and language and literary elements) to two additional criteria—vocabulary (e.g., multisyllabic words) and sentence complexity (length, embedded clauses, punctuation). Although Fountas and Pinnell (1999) mention the regularity of letter-sound spellings as a factor in determining sophistication of vocabulary, this feature is not highlighted in reports of leveled texts (e.g., Fountas & Pinnell, 1999, 2001).

#### SCALE FOR TEXT ACCESSIBILITY AND SUPPORT

The Scale for Text Accessibility and Support (STAS-1), developed by Hoffman and his associates (Hoffman et al., 2001), used a similar methodology as those used in Carver's (1976) Rauding Scale and Singer's Eyeball Estimate of Readability (SEER; Singer, 1975), in which experts use anchor passages that had been ordered according to specific criteria in leveling texts. Hoffman et al.'s system uses 5-point scales for two primary traits: decodability and predictability. *Highly decodable texts* (rated as 1) contain words with consonant–vowel–consonant (CVC) patterns, single syllables, and short high-frequency words, whereas *minimally decodable texts* (rated as 5) contain irregularly spelled words and a variety of patterns and offer little word-recognition support to the emerging reader, with three interim points of very decodable (2), decodable (3), and somewhat decodable (4). The predictability scale has a similar 5-point range, with highly predictable texts awarded a score of 1 and minimally predictable, a score of 5. The scale used four predictable features (picture support, repetition, rhyming elements, and familiar events/concepts) that texts contained to different degrees.

### *Empirical Evidence for Text-Leveling Systems*

#### GUIDED READING LEVELS

Publishers and educators have applied the text leveling of Reading Recovery and

guided reading to literally thousands of texts. Despite its widespread use, we were unable to find any reports of reliability across coders in leveling texts for either scheme. Further, although proponents of this form of leveling present it as an alternative to readability formulas, one of the only studies of its validity has reported a strong correlation between text levels and the principal factors that make up traditional readability formulas (Hatcher, 2000). Hatcher (2000) considered how five variables predicted Reading Recovery levels of 200 texts (10 at each of 20 levels) on numbers of (1) words; (2) words in the longest sentence; (3) words with six or more letters; (4) contractions, negatives, auxiliary verb plus a main verb, and auxiliary verb that changes tense; and (5) pages. Two variables—length of words and of the longest sentences—predicted Reading Recovery levels best ($R$ = .82).

We could find no studies that examined how instruction with texts ordered according to either Reading Recovery or guided reading levels influenced reading acquisition. We located a single study that examined students' reading of texts of different levels. This examination was part of Hoffman et al.'s (2001) validation of their STAS-1 ratings with Reading Recovery levels and is described shortly. We should note that studies on several of the features that figure prominently in the text-leveling scheme, particularly text predictability and illustrations, exist and have been reviewed elsewhere (Hiebert & Martin, 2001). To briefly summarize, the existing evidence suggests that overreliance on these scaffolds appears to detract from independent word recognition.

#### STAS-1

Hoffman, Sailors, and Patterson (2002) have applied the two indices that make up the STAS-1—decodability and predictability—to the first-grade texts that have been approved by the Texas Education Agency for purchase with state funds over three adoption periods: 1987, 1993, and 2000. After applying the scale to the first 1,000 words of text that beginning readers encounter in the Texas-approved programs over this time period, Hoffman et al. (2002) report that the 1987 texts were the most decodable ($\overline{X}$ = 1.2), the 2000 texts were next ($\overline{X}$ = 1.7), and the 1993 texts were least decodable ($\overline{X}$ = 2.5). On the predictability scale, the 1993 texts had the highest ratings of predictability ($\overline{X}$ = 2.5), the 2000 texts were next ($\overline{X}$ = 3.5), and the 1987 texts had the lowest ratings ($\overline{X}$ = 4.5). Analyses of cohorts of students in the state of Texas for the effects of these changes in features of predictability and decodability of their beginning reading textbooks have yet to be conducted.

However, Hoffman et al. (2001) have considered the concurrent and predictive validity of the STAS-1 in experimental contexts. With three books from each of seven levels that reflected the guided reading and Reading Recovery levels, the scale as a whole correlated at .78. Next, Hoffman et al. examined the ability of the STAS-1 and the guided reading levels to predict student accuracy and rate across three instructional conditions (preview and read, no preview, and adult modeled). Significant effects were found for condition and reader ability in the expected directions. High-ability students and those who received adult modeling had the highest performances. Because two thirds of the Hoffman et al. (2001) students were unable to read any of the texts above the criterion for accuracy (92%), it is difficult to know how well these two systems discriminate among readers of differing abilities.

### *Conclusion: Text-Leveling Systems*

The need for using expert judgment in the evaluation of text difficulty has been recognized from the initiation of work on text difficulty. As Gray and Leary's (1935) analysis showed, numerous variables cannot be evaluated quantitatively. There are many contexts in which experts' ratings according to particular criteria have been found to be highly reliable in sorting, evaluating, or judging, such as writing samples (Cooper & Odell, 1977). Hoffman et al.'s (2001) system, building on a tradition initiated by Singer (1975) and Carver (1976), illustrates how traits can be operationalized into rating schemes. Anchors can be identified and raters can be trained to code the categories with high levels of reliability. The two domains that form Hoffman et al.'s (2001) scale appear to be highly correlated, at least in the

texts that they have analyzed to date. Further, their students were either reading at the same level of accuracy across texts of all difficulty levels (96–98%) or below the specified level of accuracy on all texts (i.e., 91% or lower), making interpretations of predictive validity difficult. Further, this scale does not discriminate across texts that, at least according to other schemes, have differences in their word-recognition demands.

However, the effort of Hoffman et al. (2001) does illustrate that particular dimensions can be defined and that raters, when given clear parameters, can sort a group of texts reliably on a recognized trait of beginning reading such as decodability. The STAS-1 demonstrates that reliable teacher-based rating schemes of text difficulty can be developed. Further, classroom teachers can use the information that a particular text is highly decodable or somewhat predictable when teaching students. By contrast, the implications for teaching *Danny and the Dinosaur* and *The Cat in the Hat*—texts that differ by five guided reading levels—are not clear.

The guided reading levels fail to convey a sufficient amount of information for teachers to use in designing lessons or selecting materials that will support their students in developing proficiency in the skills that they require to read harder material. The developers of this system have failed to demonstrate the manner in which different dimensions figure into the evaluation of difficulty of text at different levels. When the scheme was limited to the very earliest stages of reading, as it was in Peterson's (1991) work, distinctions across levels may have been apparent as teachers examined books. With the extension of the system to the entire elementary period (Fountas & Pinnell, 2001), the designations of a text as level F or level J, as was illustrated in the evaluations in Table 27.1 provide little indication as to the underlying proficiencies that students require to read particular texts.

The construct of text leveling holds promise for addressing text features such as usefulness of illustrations in beginning readers' recognition of unknown words. To date, however, developers of text-leveling schemes have not followed through on this promise by providing research on how particular text features in these systems influence young children's reading at different developmental points.

## Task-Based Text Difficulty Systems

The prototypical second-trimester decodable text, *Dragons Don't Get Colds* (Raymer, 1993), is the 50th of the 75 decodable readers that make up a first-grade reading program (Adams et al., 2000). Two text-difficulty systems that illustrate task-based text-difficulty systems are applied to this book and the other four prototypes: Juel and Roper/Schneider's (1985) decodability system and Hiebert and Fisher's (2002) Critical Word Factor (CWF).

In Juel and Roper/Schneider's (1985) decodability system, an individual score from 1 to 3 is given to each word in a text. A word receives a score of 1 if it is a *transfer* word (words with regular vowel patterns such as *bag or seat*); of 2 if it is an *association* word (words with *l-* , *r-*, and *w*-controlled vowels; diphthongs; digraphs such as *law, car, boy*); and of 3 if it has irregular or unpredictable vowel patterns (words such as *come* and *pear*). The scores in Table 27.1 indicate that the decodability rating is within a narrow range, from the 1.6 of *Dragons Don't Get Colds* and *The Cat in the Hat* to the 1.9 of *Mrs. Bold*. These data indicate that the typical words in the first set of texts will be fairly evenly distributed between transfer and association words, whereas the typical word in *Mrs. Bold* will be an association word.

The second task-based text-difficulty scheme, the Critical Word Factor (CWF; Hiebert & Fisher, 2002), indicates the number of unique words per 100 running words of text that fall outside a particular curriculum. The primer curriculum, based on evaluations of tests (Menon & Hiebert, 2005), is proficiency with the 300 most frequent words and monosyllabic words with short and long vowels. As can be seen in Table 27.1, three of the texts have 5 or fewer words per 100 running words of text that fall beyond this curriculum. The fifth text—*Mrs. Bold*—has 18 unique words per 100 running words beyond this primer curriculum. If the curriculum is designated as the 100 most frequent words and monosyllabic words with CVC patterns, then the CWF would likely be higher for all of the texts. Or

if the curriculum were the 500 most frequent words and all monosyllabic words, the CWF would likely be lower for the texts.

### *Rationale for and Description of Task-Based Text-Difficulty Schemes*

As these examples show, task-based text-difficulty systems evaluate texts on their match to a curriculum, their instructional scope and sequence, or their developmental progression. The focus in this chapter is on the decodability schemes that are currently used and on alternative task-based systems such as the CWF. In that these two types of schemes have different histories, they are described separately.

*Decodability* text-difficulty schemes are of two types: a priori schemes, such as that of Juel and Roper/Schneider (1985), and instructional consistency schemes (Hoffman et al., 2002). In the former, letter–sound relationships are presented in a hierarchy of difficulty. Schemes can be more extensive than that of the three categories of Juel and Roper/Schneider, such as the eight categories of Menon and Hiebert (2005) that distinguish between words with complex consonant patterns, not simply vowel patterns. What all of these schemes have in common, however, is that any text can be reviewed against the same curriculum.

*Instructional consistency schemes* (Hoffman et al., 2002) evaluate the letter–sound relationships of a text in relation to the instructional scope and sequence of the program of which it is part. For example, if a child has been taught the /æ/ sound as in "cab," and then reads a number of /æ/ words in text (e.g. flag, rat, can), then the instructional consistency is high. In contrast, if a child encounters few /æ/ words in text, then the instructional consistency is low. Instructional consistency is usually expressed as a percentage of words that match phonics lessons. Instructional consistency formed the cornerstone of recent mandates regarding textbook purchases in the nation's two largest states (California English/Language Arts Committee, 1999; Texas Education Agency, 1997), in which particular percentages of decodable words in at least some components of first-grade programs and in first- and second-grade programs were specified: Texas, 80%; California, 90%.

The CWF is an index of two aspects of a text: (1) 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 (2) 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 (Hiebert & Fisher, 2002). Because the Text Elements by Task (TExT) software program (Hiebert & Martin, 2002) is used to identify groups of words within a text, the curriculum can be tailored for different developmental levels. Whatever the targeted curriculum of phonetically regular and high-frequency words, the CWF is an indicator of the number of words that fall outside the specified curriculum in 100 running words of text.

### *Research Validating Task-Based Text Difficulty Systems*

#### VALIDATION OF DECODABILITY SCHEMES

The most prominent of the a priori schemes has been Juel and Roper/Schneider's (1985). In the Juel and Roper/Schneider study, overall regularity ratings for basal reading texts differed only at the preprimer levels, not at the primer and first-reader levels. Descriptions of beginning reading programs from the perspective of instructional consistency have been compiled for the textbooks of most eras. The most widely publicized of these instructional consistency studies was Chall's (1967/1983) analysis of the match in the words in the texts of four basal programs relative to the instructional guidance in the accompanying teachers' editions. This paradigm was extended by Beck and McCaslin's (1978) study to include the *potential for accuracy* criterion. Beck and McCaslin's paradigm has been applied in a number of studies (e.g., Reutzel & Daines, 1987). A recent study of this type was conducted by Stein, Johnson, and Gutlohn (1999) of the texts intended for the first half of first grade from seven basal reading programs. A word had the potential for accuracy if all constituent parts could be decoded based on instruction to that point, as described in the teacher's guide for the program, or if recognition of a word by sight had been part of a lesson. Across the seven programs, Stein et al.

(1999) identified 14 components that provided three types of texts: student readers, phonics readers, or phonics support materials. One program (Scholastic's Literacy Place) had all three of the components; another program (Open Court) had only one component. Across the 14 components, the average potential-for-accuracy percentage was 59%. Two components attained Stein et al.'s criterion of 90% potential for accuracy: Open Court's student readers and Scholastic's phonics readers. Without these two components, the average percentage for potential for accuracy across texts was 53%.

Similarly, Foorman, Francis, Davidson, Harm, and Griffin (2004) examined all words in all text components (including phonics minibooks, big books, and anthologies) of six first-grade basal readers published from 1995 to 2002 using a scheme of: (1) decodable now, (2) decodable later (later instruction will make it readable), (3) holistically taught (word taught as a sight word), or (4) never decodable (neither letter–sound nor holistic information was given). When the decodable-now and holistically taught classifications were collapsed, the words in the most decodable basals were within a range of 51–85% decodable. The words in the least decodable basals ranged from approximately 25–50%.

In an earlier study, Foorman, Francis, Fletcher, Schatschneider, and Mehta (1998) considered the progress of Title I first and second graders with one of the less decodable basals and one of the most decodable basals when the first half of grade 1 relied on highly decodable texts. The instruction included differences other than the texts that students read, including different emphasis on opportunities for independent writing and spelling. Foorman et al. (1998) reported significant effects for word recognition and comprehension with the decodable texts relative to the texts of the other programs.

Precisely how the text features that Foorman et al. (1998) and Stein et al. (1999) have described influence students' reading over the long run requires substantial classroom investigations to understand. Such studies are hampered by the frequent changes that characterize programs from one copyright period to the next. For example, the Open Court program that Stein et al. (1999) and Foorman et al. (2004) identified as high in potential for accuracy has been replaced by 2000 and 2002 copyright versions. When state mandates vary in requirements from one textbook adoption to the next, the results of analyses of a publisher's program from one decade to the next can similarly vary (see, e.g., Hoffman et al., 2002).

Further, the application of instructional consistency and a priori schemes can produce quite different perspectives on the same texts. Hoffman et al. (2002) compared the results of an instructional consistency and an a priori scheme that were applied to the same texts. For instructional consistency, they used the potential-for-accuracy scores reported by the Texas Education Agency (1997) in their review. This measure is the sum of decodable words plus words taught as sight words divided by the total number of words. For the a priori scheme, they used Menon and Hiebert's (2005) eight categories to analyze each word in the same texts. The correlation between the two measures was low: $r = -.07$.

To this point, data are not available on the number of lessons that teachers need to teach for assessments of instructional consistency to be robust. However, several studies have linked a priori decodability schemes with students' success in reading in particular programs, the most widely cited of which was conducted by Juel and Roper/Schneider (1985). In this quasi-experimental study, the treatment group read from a decodable basal, and the other group from a high-frequency basal, but both groups received the same scripted phonics instruction. At the study's conclusion, groups did not differ in reading words in lists or texts from their own basals, but they did differ in decoding ability and in reading the unknown words from the other basal. The decodable group performed better on the decoding measure at interim and end-of-year assessments but not on reading words from their basal reader or a norm-referenced reading test. Juel and Roper/Schneider (1985) also examined word-level features of the two textbook programs in relation to students' performances. The decodable group was most influenced by the degree to which words were decodable, whereas the high-frequency group was most influenced by the number of times words were repeated. The conclusion of this study

was that text difficulty, as measured by decodability, was most influential during the first two trimesters of first grade.

More recently, Compton, Appleton, and Hosp (2004) have used an a priori analysis of decodability to predict students' reading performances. They found that second graders' accuracy and fluency across a 15-week period were related to the percent of high-frequency words; fluency was influenced by decodability of texts. Whereas performances of average-achieving students were influenced by the percentage of decodable words, the performances of low-achieving students were not. Compton et al. suggest that the decoding skills of low achievers may have been so poor that few words were decodable for them.

In another recent experiment, Jenkins, Peyton, Sanders, and Vadasy (2004) randomly assigned struggling readers to a tutorial with either decodable or nondecodable text. The treatments of both groups involved the same scripted tutorial lessons, only differing in the texts used for practice. Jenkins et al. used an instructional consistency criterion for decodability, with 71–84% of the words in decodable texts and 11–68% of words in the nondecodable texts consistent with their curriculum. Students in the two groups and those in a nonrandom control group performed similarly on the pretests. Although both treatment groups performed significantly higher on the posttest than the control group, the two treatment groups did not differ on any posttest measure. Jenkins et al. (2004) give two possible explanations for these patterns. First, the phonics instruction of the scripted lessons may have been sufficient for reading improvement. Second, tutors may have made texts "decodable" by directing tutees to use decoding strategies. Analyses of the books used in the nondecodable treatment (Mesmer, 2001) suggest another explanation. Many more words may have been identified as decodable in the nondecodable texts if an a priori analysis of decodability rather than an instructional consistency criterion had been applied.

### VALIDATION OF THE CWF

Hiebert (2005) examined the first-grade anthologies for the five reading programs approved by the Texas Education Agency (1997) and another mainstream textbook program not submitted to Texas. For comparative purposes, Hiebert included three historical copyrights (starting with 1962) for one of the Texas-approved programs and end-of-grade-two anthologies for all programs. Analyses showed that 41% of the unique words in current textbooks appeared once in 10 consecutive texts. Further, between 1962 and 2000, the number of unique words increased substantially, whereas word repetition was curtailed.

Another line of inquiry (Hiebert & Fisher, 2002) has considered the ability of the CWF model to predict the words that children will pause over or be unable to identify. In one study, first graders read four texts in a randomized order—two with high CWFs (a substantial number of unique words fell beyond the curriculum of the 100 most frequent words and words with CVC and long-vowel patterns) and two with low CWFs (most unique words fell within the designated curriculum). Analyses showed strong main effects for CWF on reading speed, accuracy, and comprehension, with all three variables in the direction predicted by the model.

A second set of studies has considered the effects of reading texts with different CWFs on children's reading development. In Menon and Hiebert's (2005) study, children in two classes in an inner-city school read from books that had been leveled according to a graduated CWF curriculum (i.e., the number of difficult or hard words remained consistent, but the underlying curriculum got progressively more difficult), whereas two other classes read from basal literature texts that had a consistently high CWF. Pretest scores were similar but, on the posttest, children in the low-CWF classes performed at significantly higher levels on word-list and text reading tasks than students who read from the high-CWF texts.

In a subsequent study, Hiebert and Fisher (2004) compared the performances of two groups of first-grade English-language learners who received the same scripted small-group instruction over 12 hours with those of a passive control group. One group received texts for which the CWF was 1 and the second group of texts had a CWF of 3, relative to a curriculum of the 100 most frequent words and CVC vowel patterns. Students who read from the texts with the lower

CWFs had higher fluency and accuracy levels than students who read texts with somewhat higher CWFs, and both groups had significantly higher fluency and accuracy levels than students in the passive control group.

### *Conclusion: Task-Based Schemes*

Relative to phonics schemes based on instructional consistency, a priori schemes have an advantage in representing difficulty on a clearly defined scale. Instructional-consistency schemes can be manipulated to ensure high percentages of potential for accuracy. For example, if lessons on *r*-controlled and vowel diphthongs had preceded the introduction of *Mrs. Bold* (Beck, 1993), the publishers could argue that words such as *flew, drew,* and *school* have the potential for accuracy even though the program had only a handful of phonics lessons. Unless a priori schemes are comprehensive, however, they provide little guidance for instruction. For example, the rating of 1.6 for *Dragons Don't Get Colds* (Raymer, 1993) on Juel and Roper/Schneider's (1985) scale leaves teachers with little information on which word-vowel patterns should be emphasized in lessons with struggling students.

By providing an index that is derived from a curriculum, the CWF provides teachers with an indication of what knowledge is required for students to independently read a text. This information is particularly useful in that it allows teachers to measure texts against bodies of knowledge that are viewed to be acquired developmentally. Recognition of the 100 words that appear 1,000 or more times per 1 million words of text (Zeno, Ivens, Millard, & Duvvuri, 1995) would be expected to be acquired before recognition of words that have a likelihood of appearing 100 times in a similar-sized sample or those that appear 10 times or fewer. If the curriculum is emphasizing CVC words, the word *cap* should have a higher likelihood of being recognized by children who are being taught CVC words than should words such as *cape* or *capture*.

Similar to all text-leveling systems at the current point, the CWF does not take into account the presence of highly concrete words in texts and the usefulness of background knowledge and even accompanying illustrations in children's recognition of words. Research has confirmed that children learn highly concrete words with greater ease than less concrete words (Hargis, Terhaar-Yonkers, Williams, & Reed, 1988). The inclusion of picture–text match in the guided reading levels (Fountas & Pinnell, 1996) recognizes this aspect of word learning. However, pictures can provide different levels of information, and asking children to focus on the pictures rather than on applying context strategies that integrate the use of illustrations can create problems for subsequent independent reading. One possible technique for future use that quantifies the quick usability of illustrations as a context clue has been suggested by Menon and Hiebert (2005), who evaluated the match between words that adults associated with the illustrations from pages in children's texts and the words that appeared on those pages. Additional efforts such as that of Menon and Hiebert are needed to establish how particular elements influence the difficulty of texts for beginning readers in the immediate reading task and the manner in which such elements influence proficient reading in the long run.

## Discussion: Next Steps

The most fundamental conclusion of this review is how little scholarship there has been on any of the text-difficulty schemes. We use the word *scholarship* rather than *research* because theoretical frameworks on the role of text in beginning reading, not just empirical investigations of text difficulty, are conspicuously absent. Regardless of the text-difficulty scheme, we could locate few theoretical frameworks on the role of text in beginning reading acquisition.

Gray and Leary's (1935) analysis remains the most extensive effort to identify features that may influence text difficulty. As behaviorists, they focused on readily quantifiable variables that accounted for the most variance in analyses of adult readers' performances on particular texts. Once sentence length and semantic difficulty had been identified as accounting for much of the variance, efforts to understand what these variables represented ceased for approximately four

decades. When cognitive scientists addressed the complexity of text four decades later, they offered ideas for theoretical frameworks but did not directly address texts for beginning readers.

In the current emphasis on empirical investigations, we cannot forget that empirical investigations need to build on underlying theoretical frameworks if they are to address critical questions. Regardless of the perspective on text difficulty, underlying theoretical frameworks on appropriate texts for beginning readers have either been lacking or inadequately developed. Readability formulas have emphasized two variables that can be easily quantified and that discriminate across texts. Leveling systems have included a range of variables, but the most popular of these systems has not indicated how these different variables contribute to evaluations of difficulty. Task systems have focused, at most, on a handful of word-level variables. To date, none of the systems has a comprehensive conceptual framework that differentiates the influence of variables at both the word and text levels at different developmental periods of reading. For example, the manner in which figurative and idiomatic language influences text difficulty needs consideration, even at the early levels, where many such devices can be found in children's literature. Another aspect of text for which a strong theoretical and empirical scholarship exists is the influence on reading and memory of imagery and concreteness of language (Sadoski & Paivio, 2001). At a time when children are exposed to highly visible electronic media for thousands of hours, text difficulty schemes cannot ignore the role of illustrations in texts.

Perspectives on text difficulty have been particularly lacking with respect to the kinds of texts beginning readers need *over time*. In focusing on the individual text—even when "ordered or graduated"—the readability and text-leveling schemes draw attention away from the need to consider a group of texts as the critical unit for beginning readers. More comprehensive text-difficulty schemes are needed, and these schemes need to consider progression over the entire period of reading acquisition, if students are to receive the supportive texts many require to become proficient readers.

## References


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., Hiebert, E. H., Scott, J. A., & Wilkinson, I. A. G. (1985). *Becoming a nation of readers: The report of the Commission on Reading*. Champaign, IL: Center for the Study of Reading.

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.

Beck, I. L., McKeown, M., Omanson, R., & Pople, M. (1984). Improving the comprehensibility of stories: The effects of revisions that improve coherence. *Reading Research Quarterly, 19*, 263–277.

Beck, J. (1993). *Mrs. Bold*. Auckland, New Zealand: Shortland.

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.

Carver, R. P. (1976). Measuring prose difficulty using the Rauding scale. *Reading Research Quarterly, 11*, 660–685.

Chall, J. S. (1983). *Learning to read: The great debate*. New York: McGraw-Hill. (Original work published 1967)

Chall, J. S. (1988). The beginning years. In B. L. Zakaluk & S.J. Samuels (Eds.), *Readability: Its past, present, and future* (pp. 2–13). Newark, DE: IRA.

Compton, D. L., Appleton, A. C., & Hosp, M. K. (2004). Exploring the relationship between text-leveling systems and reading accuracy and fluency in second grade students who are average and poor decoders. *Learning Disabilities Research and Practice*, *19*, 176–184.

Cooper, C. R., & Odell, L. (Eds.). (1977). *Evaluating Writing: Describing, measuring, judging*. Buffalo: State University of New York at Buffalo.

Cummings, P. (1983). *Molly's surprise*. In I. E. Aaron, D. Jackson, C. Riggs, R. G. Smith, R. J. Tierney, R. E. & Jennings (Eds.), *Scott, Foresman Reading* [Primer] (pp. 14–23). Glenview, IL: Scott, Foresman.

Davison, A., & Kantor, R. N. (1982). On the failure of readability formulas to define readable texts: A

case study from adaptations. *Reading Research Quarterly, 17*(2), 187–208.

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.

Foorman, B. R., Francis, D. J., Fletcher, J. M., Schatschneider, C., & Mehta, P. (1998). The role of instruction in learning to read: Preventing reading failure in at-risk children. *Journal of Educational Psychology, 90,* 37–55.

Fountas, I., & Pinnell, G. S. (1996). *Guided reading: Good first teaching for all children.* Portsmouth, NH: Heinemann.

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

Fountas, I. C., & Pinnell, G. S. (2001). *Guiding readers and writers: Grades 3–6.* Portsmouth, NH: Heinemann.

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

Geisel, T. S. [Dr. Seuss]. (1957). *The cat in the hat.* New York: Random House Books for Young Readers.

Good, R. H., & Kaminski, R. A. (2002). *DIBELS oral reading fluency passages for first through third grade* (Tech. Rep. No. 10). Eugene, OR: University of Oregon.

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

Gray, W. S., & Leary, B. W. (1935). *What makes a book readable.* Chicago: University of Chicago Press.

Hargis, C. H., Terhaar-Yonkers, M., Williams, P. C., & Reed, M. T. (1988). Repetition requirements for word recognition. *Journal of Reading, 31*, 320–327.

Hatcher, P. J. (2000). Predictors of Reading Recovery book levels. *Journal of Research in Reading, 23*, 67–77.

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

Hiebert, E. H., & Fisher, C. W. (2002, April). *The critical word factor in texts for beginning readers: Effects on reading speed, accuracy, and comprehension.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Hiebert, E. H., & Fisher, C. W. (2004, April). *Effects of text type on the reading acquisition of English Language learners.* Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Hiebert, E. H., & Martin, L. A. (2001). The texts of beginning reading instruction. In S. B. Neuman & D. K. Dickinson (Eds.), *Handbook of early literacy research* (pp. 361–376). New York: Guilford Press.

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

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.

Hoffman, J. V. (2002). The words in basal readers: A historical perspective from the United States. In R. Fisher, G. Brooks, & M. Lewis (Eds.), *Raising standards in literacy* (pp. 82–97). London: Routledge Falmer.

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

Jenkins, J. R., Peyton, J. A., Sanders, E. A., & Vadasy, P. F. (2004). Effects of reading decodable texts in supplementary first-grade tutoring. *Scientific Studies of Reading, 8,* 53–86.

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. L. Kamil, & P. Mosenthal (Eds.), *Handbook of reading research* (pp. 681–744). New York: Longman.

Lively, B., & Pressey, S. (1923). A method for measuring the "vocabulary burden" of textbooks. *Educational Administration and Supervision, 9*, 389–398.

McCall, W. A., & Crabbs Schroeder, L. (1979). *McCall–Crabbs Standard Test Lessons in Reading.* New York: Teachers College, Columbia University. (Original work published 1926)

Menon, S., & Hiebert, E. H. (2005). A comparison of first-graders' reading with little books or literature-based basal anthologies. *Reading Research Quarterly, 40*(1), 12–38.

Mesmer, H. A. (2001). Decodable text: A review of what we know. *Reading Research and Instruction, 40*(2), 121–141.

Michelson, J. (1985). IRA, NCTE take stand on readability formulae. *Reading Today, 2*(3), 1.

Micro Power & Light. (1999). *Readability calculations* [Computer software]. Dallas, TX: Author.

Peterson, B. (1991). Selecting books for beginning readers: Children's literature suitable for young 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.

Pikulski, J. J. (2002). *Readability.* Boston: Houghton Mifflin.

Raymer, D. (1993). *Dragons don't get colds*. Columbus, OH: Open Court SRA.

Reutzel, R., & Daines, D. (1987). The text-relatedness of seven basal reading series. *Reading Research and Instruction*, *27*(3), 26–35.

Rylant, C. (1998). *Poppleton everyday.* New York: Scholastic.

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

*Scholastic Reading Inventory: Lexile levels/performance standards*. (2002). New York: Scholastic.

School Renaissance Institute. (2000). *The ATOS readability formula for books and how it compares to other formulas*. Madison, WI: Author.

Singer, H. (1975). The SEER technique: A non-computational procedure for quickly estimating readability level. *Journal of Reading Behavior*, *7*, 255–267.

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. ED307577).

Spache, G. (1953). A new readability formula for primary-grade reading materials. *Elementary School Journal*, *55*, 410–413.

Spache, G. D. (1981). *Diagnosing and correcting reading disabilities*. Boston: Allyn & Bacon.

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.

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*. Brewster, NY: Touchstone Applied Science Associates, Inc.