# An Examination of Current Text Difficulty Indices with Early Reading Texts

**Elfrieda H. Hiebert**
TextProject, Inc.

**P. David Pearson**
University of California, Berkeley

**TextProject**

**textproject.org**

# An Examination of Current Text Difficulty Indices with Early Reading Texts

**Elfrieda H. Hiebert**
TextProject, Inc.

**P. David Pearson**
University of California, Berkeley

### Abstract

This study considers the degree to which two quantitative indices—Lexiles and Coh-Metrix—discriminate across levels of difficulty and types of beginning reading texts. The database consisted of 444 texts, representing seven text types that are part of reading/language arts instruction. These text types were distributed across seven levels of text difficulty. Analyses showed that Lexiles predicted a clear progression in difficulty across the seven levels but that these differences were due almost entirely to Mean Sentence Length (MSL), not Mean Lexical Frequency (MLF). Findings were similar for the syntax and word abstractness variables of Coh-Metrix. Of three additional Coh-Metrix variables—non-narrativity, referential cohesion, and situation model cohesion—only referential cohesion showed a progression of easier to harder across the seven text levels. Of the seven text types, trade books had the highest Lexiles, while historical textbooks had the lowest. The results of the Coh-Metrix analyses showed that all text types fell within the easy range but trade books were predicted to be the hardest and the historical textbooks the easiest. These quantitative systems validate the general order of text levels and indicate some differences in the predicted ease or difficulty of text types. The usefulness of this information—general in nature—in matching beginning readers with appropriate texts is less certain. The report concludes with an identification of next steps for supporting optimal matches of texts and beginning readers' knowledge.

# Contents

# An Examination of Current Text Difficulty Indices with Early Reading Texts

THIS STUDY CONSIDERS THE DEGREE TO WHICH CURRENTLY AVAILABLE quantitative indices discriminate across texts for beginning readers. In particular, our interest was in establishing the ability of two fairly recent text-difficulty schemes to discriminate among levels and types of early reading texts—Lexiles (Stenner, Burdick, Sanford, & Burdick, 2007) and Coh-Metrix (McNamara, Graesser, Cai, Kulikowich, & McCarthy, 2010).

Changes in perspectives on text for early reading instruction have been substantial over the past 25 years. Since these changes have been described elsewhere (e.g., Hiebert, 2005), the nature and rationale for these changes are not a focus of this report. One aspect of these changes, however, is important to note because it changed the use of quantitative indices for the creation and selection of texts in beginning reading programs—the textbook adoption guidelines of California (California English/Language Arts Committee, 1987). At that point, California (followed by Texas in 1990) stipulated that acceptable texts for its 1989 reading/language arts textbook adoption should not be manipulated to comply with readability formulas. Since that time, readability formulas have not been central to the design of reading/language arts programs (although there are indications that readability formulas never stopped being used for content-area textbook programs such as science). When California's performance in the first state-by-state comparison of the 1994 National Assessment of Educational Progress was less than stellar (Campbell, Donahue, Reese, & Phillips, 1996), both California and Texas changed their mandates for early texts from authentic literature to decodable text.

With the exception of the decodability mandates (which have not been accompanied by a valid, reliable means of establishing text difficulty), decisions about the difficulty of texts for early reading programs have been primarily qualitative. The text selections of the large-scale core reading programs appear to be based on expert judgments (presumably those of editors or authors).

Another way in which publishers have represented the difficulty of texts for beginning readers is through sorting of texts by educators. The current system, commonly called text leveling, began with Peterson's (1991) identification of features that characterized texts used in Reading Recovery. It was subsequently

applied and refined by Fountas and Pinnell's (1999) 18 guided reading levels later extended to 26 by Fountas and Pinnell (2001). The guided reading levels are differentiated along four dimensions: (a) book and print features; (b) content, themes, and ideas; (c) text structure; and (d) language and literary elements. Reports of inter-rater agreement on the sorting of texts in reading programs or, for that matter, the leveled texts of tests such as the Developmental Reading Assessment (Beaver, 1997) are not available within the archival literature or as technical reports from publishers. Consequently, it is unclear whether particular dimensions are given different weights in the sorting process at different points across the levels.

There has been no concerted effort to study either the Lexile system or Coh-Metrix indices in relation to beginning reading texts, despite the popularity of the former in the marketplace, the prevalence of the latter in research contexts, and the prominence of both in the recent publication of the Common Core State Standards (CCS; Common Core State Standards Initiative, 2010). Lexiles and the Coh-Metrix indices represent two different, although likely complementary, methods for describing quantitative features of texts that may influence readers' comprehension of them.

Lexiles are derived from the same two measures that are used to compute readability formulas: semantic difficulty, as measured by the frequency of the texts' words in a lexical database, and syntactic difficulty, as measured by sentence length. According to The Lexile Framework for Reading (Stenner et al., 2007), Lexiles range from 0 to 370 for first grade and from 340 to 500 for second grade.

Coh-Metrix is an automated tool that yields direct measures of words, sentences, and texts. Most measures of readability use indirect indices such as sentence length to predict text difficulty. By contrast, Coh-Metrix measures syntactic complexity as a function of the number of modifiers in noun phrases and the number of words before the main verb in a sentence—sentence features that have been shown to influence the comprehensibility of ideas. In work related to the CCS project (CCS, 2010), the Coh-Metrix group at the University of Memphis (McNamara et al., 2010) applied 100 measures related to word, sentence, and text features to a set of 40,000 texts. They found that eight dimensions accounted for 67% of the variability among texts. Five of these dimensions—non-narrativity, referential cohesion, situation model cohesion, syntactic complexity, and word abstractness—accounted for most of the results.

This study examines how well Lexiles and the five Coh-Metrix variables account for differences across levels and types of texts typically used for reading/language arts instruction in American K–2 classrooms. We have also included data on commonly used measures of readability: Degrees of Reading Power, which was the basis for the text difficulty on which the Coh-Metrix variables were validated (Koslin, Zeno, & Koslin, 1987), the Fry readability formula (Fry, 1968), and the Spache readability formula (Spache, 1953).

# Method

## Selection of Texts

For this first phase of investigating text difficulty of beginning reading texts, our goal was to have a set of the texts used in American reading/language arts programs that was as comprehensive as possible. In the second phase of this work, we intend to include content-area texts. But, for this first phase, we applied selected text difficulty indices to the texts that consume a substantial portion of primary students' school lives—the texts of reading/language arts blocks. Some of the texts in the reading/language arts programs are informational. However, their role is to support the reading/language arts objectives, not those of the content area. Consequently, because the present analysis aims to understand the usefulness of quantitative indices of text difficulty for early reading texts, content-area texts have not been disaggregated.

Traditionally, a distinction is made between trade books and textbooks in the publishing and marketing of books. Typically, the former are aimed at bookstores and libraries, while the latter are sold to schools. However, changes in the publishing industry, in textbook guidelines of states (e.g., California English/Language Arts Committee, 1987; Texas Education Agency, 1990), and in the marketplace have lessened the differences between trade books and textbooks. Even so, distinct text types can be identified within each group: two for trade publications and four for textbook programs. While the texts of tests are often similar to the texts of textbooks, these texts are treated as a unique text type in this analysis due to the role that tests have in schools. An excerpt from each of the seven text types appears in Table 1.

*Trade.* Texts that are bona fide trade are sometimes described as "high-quality literature." The sample of trade books for this study came from three sources: (a) Caldecott award-winning picture books, (b) picture books listed in the Read-Aloud Handbook (Trelease, 2006), and (c) the trade books on a list of grade-one literature from Accelerated Reader (Renaissance Learning, 2010). For the books on the Accelerated Reader list, presence in a public library collection was regarded as an indication that a book was of trade quality and not a textbook. Those books that appeared in the public library collection were included in the sample; those books that didn't appear weren't. The books from the other two sources were reviewed by two raters, both with teaching experience in the primary grades and knowledge of children's literature. Those books that both raters identified as appropriate for independent reading by primary-level students were included in the sample. The two raters then collaborated in sorting the books according to the text difficulty levels described below.

*Trade instructional.* This group of texts began with the publication of *The Cat in the Hat* (Geisel, 1957). Geisel used a vocabulary of 236 words (223 from a list of words that were either highly frequent in written English or were regarded to be highly familiar to young children) to produce a text demonstrating that compelling texts could be written for the learning-to-read phase. The success

### Excerpts from Each Text Type

| Text Type | Excerpt |
|---|---|
| Trade | A Duckling came out of the shell. "I am out!" he said.<br>"Me too," said the Chick.<br>"I am taking a walk," said the Duckling.<br>"Me, too," said the Chick."<br>"I am digging a hole," said the Duckling.<br>"Me too," said the Chick.<br>"I found a worm," said the Duckling. |
| Trade Instructional | Time for a bath, Biscuit.<br>Woof, woof.<br>Biscuit wants to play.<br>Time for a bath, Biscuit. Woof, woof. Biscuit wants to dig.<br>Time for a bath, Biscuit. Woof, woof. Biscuit wants to roll.<br>Time for a bath, Biscuit. Time to get nice and clean.<br>Woof, woof. In you go. Woof. |
| Textbook Core–Current | Pig in a wig is big, you see.<br>Tick, tick, tick. It is three.<br>Pig can mix. Mix it up.<br>Pig can dip. Dip it up.<br>Pig can lick. Lick it up.<br>It is six. Tick, tick, tick.<br>Pig is sad. She is sick.<br>Fix that pig. Take a sip. |
| Textbook Core–Historical | Look, Dick.<br>Dick! Dick!<br>Help Jane.<br>Go help Jane.<br>Go, Jane.<br>Go, Jane, go. |
| Text Ancillary–Decodable | Nan's Family<br>On the Mat<br>Sam sits on his mat.<br>Pam sits on Sam.<br>I am on Sam!<br>Tim sits on Pat.<br>Nan sits on Tim.<br>Tip sits on Nan. Tip. |
| Text Ancillary–Guided | Funny Faces<br>Look at the fish face.<br>Look at the fox face.<br>Look at the dog face.<br>Look at the frog face.<br>Look at the cat face.<br>Look at the flower face. Funny faces! |
| Test (GORT-4) | See Father.<br>Father is here.<br>We want to play.<br>Can you play, Mother?<br>We can play here. |

of *The Cat in the Hat* resulted in trade publishers such as Random House and Harper & Row initiating series aimed at the parent trade market. Until the late 1980s, these series were typically not used in schools other than as part of school library collections. Since that time, trade instructional texts have become part of core reading programs as well as ancillary components of reading/language arts programs. We used the texts from one of the programs available in the marketplace—the I Can Read series of HarperCollins.

Perusal of the excerpts in Table 1 suggests that there is at least a moderate amount of control in the words that have been chosen for books within this text type. One of the ways in which this control is implemented is through the presentation of a series of texts around a character—for example, Biscuit in the excerpt in Table 1. While trade instructional texts have controls on vocabulary, these controls are not as limiting as the text style commonly thought of as "Dick-and-Jane."

*Textbook core.* From a modest beginning in which a handful of graded texts provided the basis for reading instruction (e.g., the McGuffey Readers), basal or core reading programs have grown considerably. Components of reading programs at the primary grades typically include decodable readers and guided reading texts as well as songbooks, charts, CD-ROMS, DVDs, sets of library books, and workbooks. However, just as with their predecessors, current core reading programs are centered on a series of textbooks designated for each grade level that are generally called anthologies.

We used two textbook programs in this analysis: a program currently in use—Scott Foresman's *Reading Street* (Afflerbach et al., 2007)—and a historical copyright of this program—Scott, Foresman, & Company's *The New Basic Readers* (Robinson, Monroe, & Artley, 1962). We used the Scott Foresman programs for several reasons, the most prominent of which is that this is the only program still published which Chall (1967/1983) reviewed. In addition, Scott Foresman's *Reading Street* showed the greatest percentage of market share during the 2008–09 school year (Education Market Research, 2010).

*Textbook ancillaries.* Unlike other categories where texts within a category share many similarities in style, this category has at least two types that are distinctive in both word features and syntactic complexity: guided reading texts and decodables.

In 2010, guided reading texts are part of the core reading program offerings, although typically not part of the basic installation that is usually covered by state or district funds. In about one in four American first-grade classrooms (Dewitz, Jones, & Leahy, 2009), these texts form the principal reading material. A program of guided reading texts consists of individual books, 8–32 pages in length, that are clustered in levels that vary in difficulty.

The numbers of programs that fall into the textbook ancillary–guided group are many. We chose texts from a program developed in Australia (where many of these programs originated)—the program published by Wright Group

(1996). We also chose texts from a program developed by an American publisher—Ready Readers (Juel, Hiebert, & Englebretson, 1997).

Decodables are the second type of textbook ancillary. They are typically part of the basic installation of core reading programs at the beginning of the 21st century. There are also numerous sets of stand-alone programs of decodable texts. Similar to guided reading texts, the decodables are small books. Unlike the guided reading programs where the difficulty levels of books are determined on the basis of book and print features, content, text structure, and literary elements, the difficulty levels of decodables are typically a function of the phonics content represented in the texts. Those phoneme-grapheme patterns that have one-to-one correspondences (e.g., short a in cat) are typically viewed as less difficult and appear in earlier levels. Patterns where a phoneme is represented by more than one grapheme (e.g., long a in gate) are considered more difficult and come later in the sequence of texts.

Texts from two programs of decodables were used in this analysis: (a) the Open Court Reading Program (Adams et al., 2000) and (b) Reading Mastery (Englemann & Brunner, 1995).

*Tests.* The texts of tests vary considerably in their source and style, particularly in the middle to upper grades where authentic sources of texts (e.g., short stories, magazine articles) are used. Such sources are typically not used in the primary grades where, unlike textbook programs, readability formulas have continued to be used in the production of the texts. The test passages that are included in this corpus come from four sources: (a) the Developmental Reading Assessment (DRA) (Beaver, 1997)—an assessment based on a set of guided reading levels; (b) the Gray Oral Reading Test (GORT-4) (Wiederholt & Bryant, 2001); (c) two informal reading inventories—the Qualitative Reading Inventory (QRI) (Leslie & Caldwell, 2001) and the Basic Reading Inventory (BRI) (Johns, 1997); and (d) the benchmark oral reading fluency assessments of the Dynamic Indicators of Basic Essential Literacy Skills (DIBELS) (Good & Kaminski, 2002).

## Establishing Text Levels

For this initial phase, we identified seven levels of text difficulty that span the early reading period. Historically and currently, the early reading (K–2) components of core reading programs have had eight levels. Historically, basal reading programs had a reading readiness book/workbook (intended for kindergarten and/or the first month of first grade), five books for grade one (that got progressively more difficult), and two books for grade two. Currently, when a district or school purchases a core reading program, it includes a kindergarten component (that consists of a set of "little books"—decodable and/or guided in type), five texts called "anthologies" for first grade, and two for second grade. We chose seven rather than eight levels in this first round of analysis because the two levels of second grade were not reliably distinguishable. As we

state in the concluding portion of this report, the next phase of this work will examine the second-grade texts more closely.

**TABLE 2**
**Criteria for Guided Reading and DRA by Text Level**

|  | K | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| Guided Reading[1] | A | B–C | D | E | F–G | H–I | J–K | L–M |
| DRA[2] | A–2 | 3–4 | 6–7 | 8–9 | 10–12 | 14–16 | 20–24 | 25–28 |

[1] Fountas & Pinnell, 1996, 1999
[2] Developmental Reading Assessment

With the exception of trade books, all of the programs offer their texts within a scheme of text difficulty. For most of the other text types, texts were presented in a manner that made the identification of seven levels straightforward. In several cases, the number of levels in the program did not match the number identified for this study. The trade instructional category had five levels; one of the textbook ancillary–guided reading programs was based on 18 levels; and one of the tests was based on 21 levels. The ways in which the levels were clustered in the last two cases appear in Table 2. In the case of the trade instructional program, we used Scholastic's Book Wizard (a database that provides guided reading levels, DRA levels, Lexiles, and grade levels) to distribute the texts across seven rather than the original five levels designated by HarperCollins.

**TABLE 3**
**Number of Texts by Text Type and Text Level**

| Text Type | Source of Texts | # of Texts | # of Texts by Text Level | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
|  |  |  | 1 (K) | 2 (1.1) | 3 (1.2) | 4 (1.3) | 5 (1.4) | 6 (1.5) | 7 (2) |
| Trade | Various sources | 42 | 1 | 4 | 4 | 3 | 7 | 11 | 12 |
| Trade Instructional | I Can Read series | 72 | 6 | 6 | 12 | 12 | 12 | 12 | 12 |
| Textbook Core–Current | Scott Foresman (2007) | 42 | 6 | 6 | 6 | 6 | 6 | 6 | 6 |
| Textbook Core–Historical | Scott Foresman (1962) | 36 | 0 | 6 | 6 | 6 | 6 | 6 | 6 |
| Text Ancillary–Decodable | Open Court (2000), Reading Mastery (1995) | 84 | 12 | 12 | 12 | 12 | 12 | 12 | 12 |
| Text Ancillary–Guided | Ready Readers (1997), Wright Group (1996) | 84 | 12 | 12 | 12 | 12 | 12 | 12 | 12 |
| Tests | BRI, DIBELS, DRA, GORT, QRI | 84 | 5 | 5 | 4 | 16 | 17 | 20 | 17 |
| **Totals** |  | **444** | **42** | **51** | **56** | **67** | **72** | **79** | **77** |

Our aim was to have an equivalent number of texts (6) for each level of each program. As can be seen in Table 3, we could not achieve this goal for each text type. Examples of trade that fell into the lower levels were few. Most of the recommendations and award winners are texts that would be more appropriate for reading aloud to young readers, not for them to read independently. Since

the first level of the textbook core–historical consisted of a workbook with visual and auditory discrimination activities, not texts for students to read, only six difficulty levels were available for this text type.

The summary in Table 3 indicates that, if the two textbook core programs (historical and current) are clustered together, we had roughly the same number of texts of all types with the exception of the trade selections.

## Identifying Text Leveling Systems

*Readability formulas.* Efforts to quantify the difficulty of texts have been frequent since 1923 when Lively and Pressey first presented a readability formula. Readability formulas are based on an assessment of semantic difficulty (word-level) and syntactic difficulty (sentence-level). For this analysis, we provide data from three of the conventional readability formulas and a recent addition to the field that makes use of digital technology.

We chose three conventional readability formulas that each use a different index of semantic difficulty: Degrees of Reading Power (DRP), Fry, and Spache. The DRP bases its semantic index on the count of characters; the Fry on syllables per word; and the Spache against a designated list of 1,036 words that have been deemed appropriate for the primary grades. All three of these readability formulas assess syntactic complexity on the basis of words per sentence.

The fourth readability formula, Lexiles, also uses sentence length to assess syntactic complexity. However, for semantic complexity, the calculation of a text's Lexile draws on the mean frequency of the words in a text. The mean frequency of a word is derived from the rankings of words within a massive databank of well over a billion words that Metametrics has amassed over the past 25 years.

*Coh-Metrix indices.* Through an analysis of an extended database of almost 40,000 texts (K-12), McNamara et al. (2010) identified five variables (from more than 200 variables) that predicted the difficulty of texts as measured by the Degrees of Reading Power readability formula: non-narrativity, referential cohesion, situation model cohesion, syntax, and word abstractness. McNamara et al. have suggested that data on the variables be presented as percentiles and in a consistent manner as illustrated in Figure 1 with data on the five dimensions for *Morris Goes to School* (Wiseman, 1983). We refer to data on this exemplar in defining the variables.

1. *Non-narrativity.* Narrative text tells a story, with characters, events, places, and things that are familiar to the reader and is closely affiliated with everyday oral conversation. Texts that follow a narrative structure have low percentiles on this scale. *Morris Goes to School,* with a percentile of 13 on this measure, falls on the easy or highly comprehensible end of this scale.

2. *Referential cohesion.* High cohesion texts contain words and ideas that overlap across sentences and the entire text, forming threads that explicitly connect the text elements for the reader. Similar to non-narrativity, a high

**FIGURE 1**

**Coh-Metrix Variables as Percentiles for *Morris Goes to School***
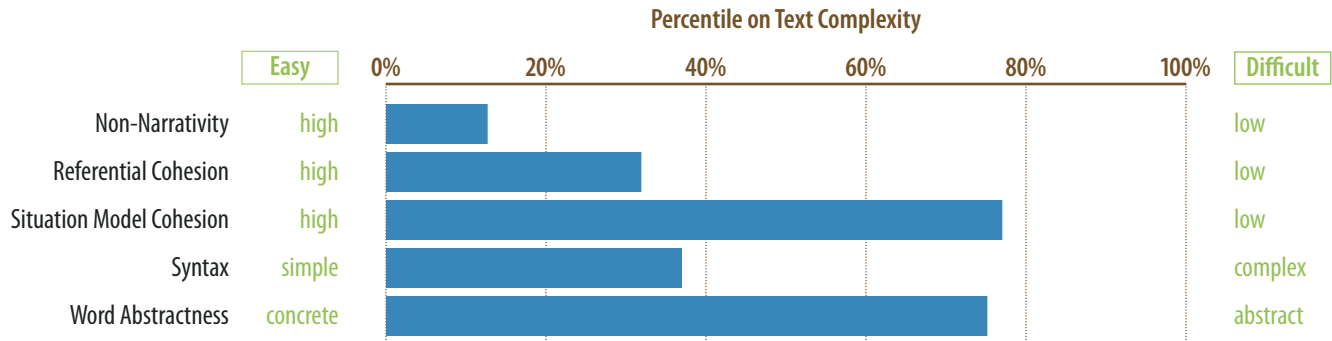


percentile on referential cohesion indicates that a text is difficult and has few of the threads that support explicitness for readers. With a percentile of 32 on this measure, *Morris Goes to School* falls on the easy half of the scale.

3. *Situation model cohesion.* Causal, intentional, and temporal connectives help the reader to form a more coherent and deeper understanding of the text. A high percentile on situation model cohesion means lower levels of this feature and, consequently, more obstacles for comprehension for readers. Thus, a high percentile on this variable indicates a more difficult text. On this particular variable, the percentile of 77 places *Morris Goes to School* on the difficult half of the scale.

4. *Syntactic simplicity.* Sentences with few words and simple, familiar syntactic structures are easier to process and understand. When texts have high percentiles on this dimension, they have complex syntactic structures, which suggest that processing will also be complex. The percentile of 37 for syntax means that *Morris Goes to School* is relatively easy on this measure.

5. *Word concreteness.* Concrete words evoke mental images and are more meaningful to the reader than abstract words. High percentiles on this dimension mean that texts have a substantial number of abstract words. Higher portions of abstract words, in turn, make texts more difficult to comprehend. With a percentile of 75, *Morris Goes to School* is judged to have a substantial number of abstract words that could impede comprehension.

The Coh-Metrix analysis also provided data on two variables that, while not prominent in the upper-grade analysis, may be important for beginning readers: the familiarity of words and the number of unique words in relation to total words in a text (i.e., type-token ratio).

# Results

## Conventional and Current Readability Formulas

*Text levels.* With respect to the readability data for the text levels as represented in Table 4, the only readability formula that shows a clear progression across the seven levels was Lexiles. The means for the two variables that contribute to a Lexile score are provided in Table 4: Mean Sentence Length (MSL) and Mean Lexical Frequency (MLF). An examination of the progressions for the two variables shows that only MSL shows a steady progression from one level to the next. The means for the other variable—MLF—show limited variation from one level to the next. All of the means are within a range of 3.6–3.8—a limited range for vocabulary. Correlations for the two variables relative to the levels also show that this progression is a reflection of differences in sentence length and not of vocabulary: .57 for MSL and .06 for MLF.

**TABLE 4**

**Readability Measures (Means) by Text Levels**

| Text Levels | DRP | Fry | Spache | Lexile | MSL[1] | MLF[1] |
|---|---|---|---|---|---|---|
| 1 | 1.6 | 1.3 | 1.9 | 86.9 | 4.9 | 3.8 |
| 2 | 1.6 | 1.1 | 1.8 | 140.0 | 5.0 | 3.6 |
| 3 | 1.6 | 1.1 | 1.8 | 238.0 | 6.1 | 3.7 |
| 4 | 1.6 | 1.3 | 1.8 | 238.2 | 6.4 | 3.8 |
| 5 | 1.8 | 1.6 | 2.0 | 346.0 | 7.2 | 3.7 |
| 6 | 2.0 | 2.0 | 2.2 | 420.6 | 8.0 | 3.7 |
| 7 | 2.2 | 2.6 | 2.3 | 489.1 | 8.8 | 3.7 |

[1] MSL (mean sentence length) and MLF (mean lexical frequency) were provided as part of the Lexile analysis of the texts. Although they are not defined as readability measures, they are included here as supplementary information.

The other three readability formulas also correlated highly with text level. However, the Fry and Spache yielded higher levels of difficulty for the pre-Kindergarten compared to the first level of first grade. All three indices (DRP, Fry, and Spache) showed very little or no differences in text difficulty for the first four levels of text.

*Text types.* The trade texts had the highest readability indices (i.e. most difficult) of all seven text types on all four measures (see Table 5). Based on these four readability indices, the trade texts were substantially more difficult than the other six text types. It should be noted that mean sentence length is substantially higher for the trade selections—9.2 compared to an average of 6.6 for the other six text types.

The texts from the textbook core–historical were rated the least difficult on all four measures. While the most difficult set of texts (trade) had the longest sentences, the easiest set of texts (textbook core–historical) had the shortest sentences. The remaining five text types, ranked in decreasing level of difficulty,

were: textbook core–current, tests, trade instructional, text ancillary–decodable, and text ancillary–guided. These ranks were relatively consistent on the four readability indices with the exception of text ancillary–decodable, which varied widely from index to index.

**TABLE 5**
**Conventional and Current Readability Indices for Text Types**

| Text Types | DRP | Fry | Spache | Lexile | MSL[1] | MLF[1] |
|---|---|---|---|---|---|---|
| Trade | 2.4 | 2.8 | 2.5 | 534.6 | 9.2 | 3.7 |
| Trade Instructional | 1.8 | 1.6 | 1.9 | 276.0 | 6.4 | 3.7 |
| Textbook Core–Current | 1.9 | 1.7 | 2.0 | 320.7 | 6.6 | 3.6 |
| Textbook Core–Historical | 1.6 | 1.3 | 1.5 | 185.8 | 5.9 | 3.7 |
| Text Ancillary–Decodable | 1.6 | 1.3 | 2.0 | 315.7 | 6.9 | 3.7 |
| Text Ancillary–Guided | 1.8 | 1.5 | 1.9 | 228.4 | 6.2 | 3.7 |
| Tests | 1.8 | 1.8 | 1.9 | 333.2 | 7.5 | 3.8 |

[1] MSL (mean sentence length) and MLF (mean lexical frequency) were provided as part of the Lexile analysis of the texts. Although they are not defined as readability measures, they are included here as supplementary information.

## Coh-Metrix Indices

*Text levels.* The means for the five variables, plus word familiarity and type-token ratio, are presented in Table 6 for the seven levels of text. On the first dimension—non-narrativity—the indices for all text levels are low with no clear distinctions among the text levels. This indicates that elements of narrative are easily identifiable in all levels of text. This pattern is not unexpected in that the texts are designed for beginning reading/language arts instruction.

**TABLE 6**
**Means for Coh-Metrix Indices by Text Level**

| Text Levels | Non-narrativity | Referential cohesion | Syntactic complexity | Word abstractness | Situation model cohesion | Familiarity | Type/ Token |
|---|---|---|---|---|---|---|---|
| 1 | 20.9 | 9.3 | 4.4 | 35.2 | 78.5 | 1.9 | .6 |
| 2 | 18.9 | 10.5 | 10.7 | 34.9 | 79.9 | 2.3 | .5 |
| 3 | 19.8 | 14.6 | 7.3 | 42.7 | 62.7 | 2.2 | .5 |
| 4 | 14.6 | 20.5 | 7.7 | 45.8 | 64.7 | 2.1 | .5 |
| 5 | 17.5 | 32.0 | 10.2 | 37.2 | 54.7 | 2.2 | .5 |
| 6 | 19.7 | 39.8 | 12.7 | 37.4 | 52.6 | 2.2 | .6 |
| 7 | 18.6 | 46.2 | 16.5 | 37.4 | 53.4 | 2.2 | .5 |

Referential cohesion is the only variable of the five that progresses from easier to harder across the seven levels of text. This progression means that ideas and vocabulary are more cohesive at the beginning levels than at the higher levels of text.

As was the case with the readability formulas, syntactic complexity ranks the text levels in the expected order (with one reversal). From text level 1, where the percentile is 4.41, to text level 7, where the percentile is 16.54, the increase in syntactic complexity is steady. There is one exception—the relatively high level of syntactic complexity for text level 2. Further analyses with a substantially larger dataset are needed to determine if this shift represents a particular type of text used at the point where students are expected to begin reading independently or whether this pattern is an artifact of particular texts in the database.

Word abstractness had a relatively restricted range indicating that the texts did not differ very much on this construct. There was an increase in abstractness over text levels 2 through 4 but abstractness fell with level 5 and stayed flat for the remaining two levels where texts would be expected to be most abstract.

To some extent, this pattern is also evident in the familiarity index (see column 7 of Table 6). Both word abstractness and familiarity would be expected to shift across the seven text levels, with movement toward more complex vocabulary. Further analyses are needed to determine if these patterns reflect the decisions underlying data analysis within the Coh-Metrix system. However, it should be noted that MLF (part of the Lexile data in Table 5) also shows little variation across levels of texts.

For situation model cohesion, scores go in the opposite direction from that predicted by the model. That is, the texts of the earlier levels are more difficult than those of the later levels. This pattern requires further investigation. It may be that beginning texts do not give students the causal and temporal links that support comprehension. It may also be that the level of complexity in these early reading texts is sufficiently low that such links are not appropriate.

The final piece of data provided by the Coh-Metrix analysis was the type-to-ken ratio—the number of different words relative to total number of words. Surprisingly, the type-token ratio stayed fairly consistent across the levels of text. Type-token ratio is, and has been, considered a critical design element in texts for beginning readers. Consequently, the ratio was expected to be high in the pre-Kindergarten texts and gradually decrease over the remaining text levels. In the current analysis, this expected progression in type-token ratio was not found.

*Text types.* The Coh-Metrix indices (Table 7) did distinguish between the two text types that were the most distinctive within this group, specifically the trade and the textbook core–historical texts. The differences were in the anticipated direction. The trade texts were written to entertain, teach about concepts or provide information, not to yield texts that support development of particular reading skills. The textbook core–historical texts were written according to a formula that specified which words could be included, the sequence in which words were introduced, and the number of repetitions of words.

The texts classified as trade had high scores on all of the indices, indicating that these texts were relatively more difficult than other texts in this sample. The textbook core–historical set of texts had extremely low scores on the seven indices, indicating that these texts were very easy.

**TABLE 7**

**Means for Coh-Metrix Indices by Text Type**

| Text Types | Non-narrativity | Referential cohesion | Syntactic complexity | Word abstractness | Situation model cohesion | Familiarity | Type/ Token |
|---|---|---|---|---|---|---|---|
| Trade | 42.8 | 22.9 | 21.7 | 44.8 | 19.5 | 2.3 | .6 |
| Trade Instructional | 25.4 | 11.2 | 17.3 | 37.8 | 6.0 | 2.1 | .5 |
| Textbook Core–Current | 45.9 | 7.9 | 32.5 | 34.0 | 5.5 | 2.3 | .5 |
| Textbook Core–Historical | 12.1 | 2.3 | 11.3 | 19.3 | 5.3 | 1.8 | .4 |
| Text Ancillary–Decodable | 38.0 | 7.8 | 16.4 | 17.7 | 12.7 | 2.1 | .5 |
| Text Ancillary–Guided | 43.1 | 12.2 | 16.9 | 18.6 | 8.1 | 2.3 | .5 |
| Tests | 22.1 | 28.5 | 17.1 | 27.9 | 14.9 | 2.1 | .6 |

## Summary and Conclusions

All four readability formulas generally showed increasing levels of difficulty with higher levels of text. Of the four, the Lexile index increased for each level of text although the increase from level 3 to level 4 was very small. The other three indices (DRP, Fry, and Spache) also trended in the expected direction but all three yielded relatively flat results for the lowest three levels of text. The Fry and the Spache indices assessed level 1 text (pre-kindergarten) as slightly more difficult than level 2 text. The analyses showed that it was sentence length and not mean lexical frequency that accounted for the predictive strength of Lexiles. As expected, the complexity of sentences influences comprehensibility of texts for beginning readers. However, manipulating sentences to make them less complex does little to increase readability for students in the early phases of learning to read (Brennan, Bridge, & Winograd, 1986). At the very earliest stages of reading, word frequency and patterns appear to be the critical variables, not syntactic complexity.

The Coh-Metrix constructs showed some association with text levels. Referential cohesion increased consistently across the text levels. Like the Lexile result, referential cohesion increased very little from text level 3 to text level 4. Syntactic complexity also trended in the expected direction with only one reversal. Non-narrativity, word abstractness and situation model cohesion did not predict text levels, at least not in this sample of texts for beginning readers. Situation model cohesion generally decreased over the text levels.

According to this unexpected result, the texts were generally more difficult at the pre-kindergarten level and were less difficult through grades one and two.

When texts were grouped by type, both the readability indices and the Coh-Metrix variables showed substantial variation from type to type. All of the indices but one identified the textbook core–historical as the least difficult of the types. However, there was little consistency among the indices for the remaining text types. This is not surprising in that the offerings for early reading instruction have always been numerous and diverse (see, e.g., Aukerman, 1984). In this first phase of our investigation, we intentionally selected texts so that the range of materials would be covered.

This variation, however, raises potential issues in identifying a text difficulty system that can be applied to the range of texts found in early reading instruction. Spache, which was designed for a specific type of text, performed perfectly within that text type. Further, guided reading texts have their own rationale and criteria for assigning text difficulty. Does this imply that several different types of difficulty systems are necessary—one for each type of text?

Before we respond to this question, a clarification between text type and genre seems in order. In this study, we examined different types of texts, most of which fall within the same genre—narrative texts. We intentionally did not include texts of science and social studies—texts intended to communicate information. We deliberately chose not to include informational texts because, as others have argued (Chall, Bissex, Conard, Harris-Sharples, 1996; Duke, Bennett-Armistead, & Roberts, 2003), the criteria for establishing the difficulty of informational texts may be different from those for narrative texts. Chall et al. (1996) went so far as to distinguish text difficulty scales for four genres of content-area texts: life sciences, physical sciences, narrative social studies, and expository social studies.

The question of multiple schemes for different types of texts that are used for beginning reading instruction is a different matter. At present, we have such a system. Different competing models are offered, each with a different method for establishing text difficulty—readability formulas for high-frequency words, qualitative levels for guided reading, and indices of decodability for decodable texts. Since our linguistic system (and the act of reading itself) has many facets, no single text type is likely to support students' development of the desired set of reading skills. Hiebert (1999) has suggested that single-criterion texts (i.e., texts that emphasize phonetically regular words; texts that emphasize high-frequency words; and texts that emphasize highly concrete words) may be needed—at least to the point where explicit integration of reading skills is appropriate. Texts that exemplify the extremes of a genre—especially when used as the core of a beginning reading program—may not support the full development of readers. We believe that a comprehensive model of texts for beginning readers and a complementary text difficulty scheme has yet to receive the attention that this topic deserves.

## Next Steps

We believe that there are some clear-cut next steps in the development of texts for early reading instruction. These texts must provide appropriate—and increasing—levels of complexity as students begin to learn to read and gradually increase their skills and competencies. We must find more and better ways to characterize text complexity and its role in learning to read. Deeper understanding of text complexity (and its measurement) offers the possibility of meeting the needs of beginning readers more effectively. In particular, contemporary digital resources offer ways of describing and identifying texts that were not available to researchers only a few short years ago. Quantitative indices may never provide the complete description; however, the new possibilities they bring to bear on the problem may result in more effective instruction for large numbers of children. An extraordinary wealth of information on linguistic corpora has emerged over the last decade. Researchers now have the opportunity to test hypotheses on large digital databases of texts. Knowledge gained in this new environment may lead to better selection (and construction) of texts for early reading instruction. With this vision in mind, we offer the following recommendations for next steps.

### • A Substantially Expanded Database

The database of texts needs to be expanded in several ways. From the present analysis, we saw that difficulty indices varied among text types. A larger database (using the same text types) would allow the disaggregation of text type and text level. This would be a relatively easy step. However, the database should be expanded to include other text genres. From the outset of this project, we have recognized that issues of text difficulty may be quite different for informational texts compared to narrative texts. It may be especially important to include science texts in the database. Preliminary work with publishers has identified a list of K–2 science texts that are available in digital form.

Many early reading researchers regard second grade as a critical period for consolidation of basic reading proficiencies and development of vocabulary for higher-level reading. The database for grade two should be expanded to include comprehensive sets of narrative and informational texts and, thereby, allow more sophisticated exploration of this critical period in early reading development. Since much of the work on text difficulty has been carried out on an extensive database of texts used in grades 3 through 12, an expansion of the grade 2 database would allow much needed comparisons of results from the "reading to learn" arena (i.e., grades 3 through 12 and beyond) and the "learning to read" arena (i.e., K–2).

In expanding the database, it is essential that selection of texts be carried out with great care. Simply using large available databases without regard for the distribution and representativeness of the texts within these samples is unlikely to be productive. Hypotheses about the usefulness of particular indices in rela-

tion to particular text types and genres need to be tested. It may also be that particular text features are influential at some levels but not at others.

### • *Analyses of Larger Units of Text Rather than Single Texts*

Some text characteristics are defined for, or take on added meaning in the context of, larger units of text. The type-token ratio, for example, is a key design consideration in beginning reading texts. This ratio is relevant for single texts, but it is also important to consider the type-token ratio calculated over the set of texts that a beginning reader might encounter in a slightly longer time frame. That is, the type-token ratio for the texts encountered over a week may be more important than the type-token ratio of any individual text. Since individual texts are very short in beginning reading programs, the unit of text on which type-token ratios (and other text characteristics) are calculated is an important consideration. Having an expanded database of beginning reading texts would allow examination of text characteristics defined on larger units of text. For example, all six of the individual texts that comprise the first level of the textbook ancillary–guided texts would be treated as a single text. If a program (i.e. a series of texts) is designed to teach children to read, it should include, at the very least, a modicum of repetition across a core group of words. Analyses of expanded databases using larger units would allow researchers to explore variations in repetition (and other constructs) empirically.

### • *An Early-Reading Specific Framework for Text Difficulty*

A more ambitious next step is to conduct analyses based on frameworks specific to early reading (see, e.g., Cunningham, Spadorcia, Erickson, Koppenhaver, Sturm, & Yoder, 2005). In addition to type-token ratios, there are several high priority candidates for analyses in early reading texts. We need to know more about the distributions of high frequency words, phonetically regular words, morphological derivatives, and highly concrete, imagable words and their interrelationships in various types of early reading texts.

### • *Validating Models with Students*

Ultimately, however, no matter how extensive the digital database, a text difficulty system needs to be validated with data on readers' performances. Are the texts of one level easier to read than the texts of a subsequent level for young readers at a specified developmental point? A text difficulty system for beginning reading is only as good as its ability to identify what makes one text hard and another one easy. Expanding the database, refining a text difficulty system for beginning reading, and empirically testing the resulting system (or systems) with beginning readers would require an allocation of scarce resources. The costs of not addressing the opportunity, however, are surely greater by several orders of magnitude.

# References

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

Afflerbach, P., Blachowicz, C. L. Z., Boyd, C. D., Cheyney, W., Juel, C., Kame'enui, E.J., et al. (2007). *Reading street.* Glenview, IL: Scott Foresman.

Aukerman, R. C. (1984). *Approaches to beginning reading.* New York, NY: John Wiley & Sons.

Beaver, J. (1997). *Developmental reading assessment.* Glenview, IL: Celebration Press.

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

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

Campbell, J. R., Donahue, P. L., Reese, C. M., & Phillips, G. W. (1996). *NAEP 1994 reading report card for the nation and the states: Findings from the National Assessment of Educational Progress and trial state assessments.* Washington, DC: National Center for Education Statistics.

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

Chall, J. S., Bissex, G. L., Conard, S. S., & Harris-Sharples, S. H. (1996). *Qualitative assessment of text difficulty: A practical guide for teachers and writers.* Cambridge, MA: Brookline, MA.

Common Core State Standards Initiative (2010). *Common Core State Standards for English Language Arts & Literacy in History/Social Studies, Science, and Technical Subjects.* Washington, DC: CCSSO & National Governors Association.

Cunningham, J. W., Spadorcia, S. A., Erickson, K. A., Koppenhaver, D. A., Sturm, J. M., & Yoder, D. E. (2005). Investigating the instructional supportiveness of leveled texts. *Reading Research Quarterly, 40* (4), 410–449.

Dewitz, P., Jones, J., & Leahy, S. (2009). Comprehension strategy instruction in core reading programs. *Reading Research Quarterly, 44* (2), 102–126.

Duke, N. K., Bennett-Armistead, V. S., & Roberts, E. M. (2003). Filling the great void: Why we should bring nonfiction into the early-grade classroom. *American Educator, 27* (1), 30–35.

Education Market Research (2010). EMR research corner: Reading market: 2010. In *The complete K–12 newsletter. Retrieved February 15, 2010,* http://educationmarketresearch.com

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

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

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

Fry, E. B. (1968). A readability formula that saves time. *Journal of Reading, 11,* 513–516, 575–578.

Geisel, T. (1957). *The cat in the hat.* New York, NY: Random House.

Good, R. H., & Kaminski, R. A. (Eds.). (2002). *Dynamic indicators of early literacy skills* (6th Ed.). Eugene, OR: Institute for the Development of Educational Achievement.

Hiebert, E. H. (1999). Text matters in learning to read (Distinguished Educators Series). *The Reading Teacher, 52,* 552–568.

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

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

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

Koslin, B. I., Zeno, S., & Koslin, S. (1987). *The DRP: An effective measure in reading.* Brewster, NY: TASA.

Leslie, L., & Caldwell, J. (2001). *Qualitative reading inventory—3.* New York, NY: Addison Wesley Longman.

Lively, B., & Pressey, S. (1923). A method for measuring the "vocabulary burden" of textbooks. *Educational Administration and Supervision, 99,* 389–398.

McNamara, D., Graesser, A., Cai, Z., Kulikowich, J., & McCarthy, P. (February 2010). *Coh-Metrix* (Report to the Gates Foundation/Text Complexity Project). Memphis, TN: University of Memphis.

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.

Robinson, H.M., Monroe, M., & Artley, A.S. (1962). *The new basic readers.* Chicago, IL: Scott, Foresman & Company.

Spache, G. (1953). A new readability formula for primary-grade reading materials. *The Elementary School Journal, 53* (7), 410–413.

Stenner, A. J., Burdick, H., Sanford, E. E., & Burdick, D. S. (2007). *The Lexile framework for reading* (Technical report). Durham, NC: Metametrics.

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

Trelease, J. (2006). *The read-aloud handbook.* New York, NY: Penguin.

Wiederholt, J.L. & Bryant, B.R. (2001). *Gray oral reading test (GORT-4)* (4th Ed). Austin, TX: PRO-ED.

Wiseman, B. (1983). *Morris goes to school.* New York: NY: HarperCollins.

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

TextProject, Inc. is a non-profit public benefit corporation. Its aim is to bring beginning and struggling readers (of any age) to high levels of literacy through a variety of strategies and tools, particularly the texts used for reading instruction.

Find out more at **textproject.org**