



READING RESEARCH REPORT

#11.03

June 2011

Using Multiple Sources of Information in Establishing Text Complexity

Elfrieda H. Hiebert

TextProject, Inc.

University of California, Santa Cruz





© 2011 TextProject, Inc. Some rights reserved.



This work is licensed under the Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 United States License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/3.0/us/> or send a letter to Creative Commons, 171 Second Street, Suite 300, San Francisco, California, 94105, USA.

“TextProject” and the TextProject logo are trademarks of TextProject, Inc.

Cover photo © istockphoto.com/aldomurillo. All rights reserved. Used under license.

RRR-11.03 V.1.00 JUNE 2011

Using Multiple Sources of Information in Establishing Text Complexity

Elfrieda H. Hiebert
TextProject, Inc.
University of California, Santa Cruz

Abstract

A focus of the Common Core State Standards/English Language Arts (CCSS/ELA) is that students become increasingly more capable with complex text over their school careers. This focus has redirected attention to the measurement of text complexity. Although CCSS/ELA suggests multiple criteria for this task, the standards offer a single measure of text complexity—Lexiles. In this paper, I propose that additional quantitative measures are available—including the two components of a Lexile rating—that can provide more comprehensive views of text complexity. I apply these two “intra-Lexile” measures and one additional measure, referential cohesion, to sets of exemplars for grade bands 2–3 and 4–5 within the CCSS/ELA. The analyses suggest that conclusions about text complexity vary considerably when multiple quantitative measures are used, rather than a single, omnibus index.

READING RESEARCH REPORT #11.03

June 2011

TextProject, Inc.

SANTA CRUZ, CALIFORNIA

Contents

Introduction	1
TABLE 1 Original and Recalibrated Lexile Ranges for CCSS/ELA Grade Bands ¹	2
A Short History of Readability Formulas	3
Traditional readability formulas	3
A new generation of readability formulas: Lexiles	3
Critiques of readability formulas	4
Identifying a Set of Quantitative Measures of Text Complexity	5
Intra-Lexile Measures: MSL and MLWF	5
TABLE 2 Typical Ranges for Word Frequency and Sentence Length	6
Referential cohesion	6
TABLE 3 Typical Averages for Referential Cohesion (Argument and Stem)	8
Applying the Quantitative Measures to Sets of Texts	8
Grade 2–3 exemplar texts	8
TABLE 4 Quantitative Indices: Exemplar Texts from CCSS Grade Band 2–3	9
Grade band 4-5	10
TABLE 5 Quantitative Indices: Exemplar Texts from CCSS Grade Band 4–5	10
TABLE 6 Relationships among Measures	10
Using Our Professional Expertise and Experience	11
References	13

Using Multiple Sources of Information in Establishing Text Complexity

“Every-single-day,” I told him for the second time this week. For the twentieth time this month. The hundredth time this year? And the past few years? “And did Papa sing, too?”

(*Sarah: Plain & Tall*, MacLachlan, 1985)

Dyeing is staining fabric with colors. Fabric is dipped in a colored liquid. The liquid is called a dye. Some dyes are made from plants. The dyed fabric can be used to make clothing.

(*Art Around the World*, Leonard, 1998)

HOW DIFFICULT ARE THE TEXTS FROM WHICH THESE TWO EXCERPTS COME? At what point in their school careers should students be expected to read these complex texts with understanding? Answers to such questions always have been central to educators’ efforts to match readers with appropriate texts. Answers to such questions are even more in the foreground currently because of the emphasis placed on text complexity in the Common Core State Standards (CCSS)/English Language Arts (ELA) (CCSS Initiative, 2010). Indeed, the CCSS/ELA’s inclusion of a standard on text complexity (Standard 10) represents the first time that a standards document, either at a state or national level, has focused on this issue.

Text complexity, as defined in the CCSS/ELA (2010), is the “inherent difficulty of reading and comprehending a text combined with consideration of reader variables (Glossary, p. 43). Determining text complexity involves qualitative components (e.g., levels of meaning, structure, knowledge demands), quantitative components (e.g., readability measures and other scores of text complexity), and reader-task components (e.g., reader variables such as motivation, knowledge, and experiences; task variables such as purpose and questions). Such a tripartite system for characterizing text complexity fits with what is known about texts and readers. As Gray and Leary (1935) demonstrated many decades ago, numerous text features can influence comprehension. Further, differences in reader characteristics as well as the purposes, uses, and contexts of reading can mean that a single text varies considerably in its comprehensibility, even for the same reader (Rand Reading Study Group, 2002).

TABLE 1
Original and Recalibrated Lexile Ranges for CCSS/ELA Grade Bands¹

Text Complexity Grade Band	Original Lexile Ranges	Recalibrated Lexile Ranges
K-1	N/A	N/A
2-3	450-725	450-790
4-5	645-845	770-980
6-8	860-1010	955-1155
9-10	960-1115	1080-1305
11-CCR	1070-1220	1215-1355

¹ Adapted from *Common Core State Standards*, Appendix A, p. 8.

Although the CCSS/ELA tripartite system of establishing text complexity is well reasoned and reasonable, not all of the system’s components had been operationalized when the standards were released. In its final form, the CCSS/ELA gives explicit guidance for determining only the quantitative component and, even for that component, it describes only a single measurement scheme—Lexiles, a recent form of readability formula (Smith, Stenner, Horabin, & Smith, 1989). Further, the Lexiles it describes have been recalibrated from longstanding recommendations of Lexiles for particular grade levels to a “grade-by-grade ‘staircase’ from beginning reading to the college and career readiness level” (CCSS Initiative, 2010, p. 8). Beginning with the grade 2–3 band, text-complexity levels have been increased to ensure text levels of college and career by the end of high school (see Table 1). The explicit parameters for Lexiles by grade bands, the ease of obtaining Lexile scores, and the lack of ready access to validated qualitative rubrics mean that policy-makers and educators could place considerable weight on Lexiles in choosing texts for instruction and assessment.

Quantitative information about a text, including Lexiles, can be useful in getting a general sense of a text’s difficulty, especially from among many texts. At the same time, any quantitative information requires interpretation. Professionals such as teachers or doctors know that basing their decisions on data from several different quantitative measures is preferable to relying on a single number or piece of information. Relying on a single data point can lead to unintended consequences.

In this paper, I identify several types of quantitative data that can be brought to bear on the evaluation of the complexity of texts. To demonstrate how these pieces of quantitative information can be used in tandem, I apply them to texts that were identified within the CCSS/ELA as exemplars of grade-appropriate texts. While the paper’s focus is on the use and interpretation of quantitative data, I stress that the use of these data is only the first step in evaluating text complexity. Once quantitative data establish that particular texts that are “within the ballpark,” the hard work of qualitatively analyzing the demands of

texts in relation to different readers and tasks begins. Before I discuss these additional sources of quantitative data, however, I offer a brief review of what we know about readability formulas.

A Short History of Readability Formulas

Readability formulas have had almost a century of use in American reading instruction. During this time, reading educators have learned a great deal about their uses (and also potential abuses and misuses).

Traditional readability formulas

Beginning with Lively and Pressey (1923), researchers have proposed more than 200 readability formulas (Klare, 1984). Almost without exception, readability in these formulas is based on syntactic and semantic complexity. Typically, the number of words per sentence determines syntactic complexity. Semantic complexity is measured by either word familiarity as defined by inclusion on a word list or the number of syllables per word.

From the 1920s through the 1980s, readability formulas were viewed to be so definitive that syntactic and semantic features were manipulated to produce texts with specific readability levels (Green & Davison, 1988). As cognitive psychology perspectives became prominent in the 1970s and 1980s, researchers reported that such manipulations could hinder rather than facilitate comprehension. Comprehension was higher on texts with precise language and coherent structures—and higher readability levels—than easier texts with less specific vocabulary and less coherence (e.g., Beck, McKeown, Omanson, & Pople, 1984).

Critiques of readability formulas in creating or manipulating texts were communicated in *Becoming a Nation of Readers* (Anderson, Hiebert, Scott, & Wilkinson, 1985), a message that struck a chord with teachers. By 1990, the two largest U.S. states that also have state-approved lists for textbook purchases, California and Texas, had mandated that reading textbooks needed to consist of authentic literature (California English/Language Arts Committee, 1987; Texas Education Agency, 1990). Even when mandates for decodable texts replaced those for authentic literature in the early 2000s, publishers were not required to provide evidence of texts' readability.

A new generation of readability formulas: Lexiles

At the same time that reading researchers were describing the limitations of readability formulas, several projects were under way to conduct readability formulas digitally. The most prominent of these efforts has been the Lexile Scale (Smith et al., 1989). Like its predecessors, Lexiles are based on a mathematical algorithm of syntactic and semantic measures. The syntactic measure is straightforward: the mean sentence length (MSL) of a sample of sentences. The semantic component—the Mean Log Word Frequency (MLWF)—is based

on a word's relative frequency to other words in a databank that began with the 5 million words in Carroll, Davies, and Richman's (1971) analysis of grade three through nine schoolbooks of the 1960s. The databank has grown to well over a billion total words (A. J. Stenner, April 15, 2010). The number of unique words or *types* within the databank is less certain but it undoubtedly numbers much more than the 86,741 unique words that Carroll et al. identified. The MSL and MLWF are then entered into the formula that produces a Lexile on a scale that spans 0 (easiest texts) to 2000 (most complex texts).

Critiques of readability formulas

Criticisms of readability formulas have been raised since their inception and these apply to the digital generation of readability formulas such as Lexiles as well. One problem was raised earlier. Specifically, short sentences and frequent words that result in an “easy” designation of text complexity do not necessarily support high levels of comprehension.

A second criticism has to do with the potential inflation of informational text difficulty and the potential deflation of narrative text difficulty. Because informational texts use precise and often rare vocabulary, rare words are often repeated (Cohen & Steinberg, 1983). Readability formulas fail to treat this rare but frequent vocabulary differently, despite evidence that readers become more facile with vocabulary after several repetitions (Finn, 1978). In narrative texts with substantial amounts of dialogue, average sentence length can be influenced in that people typically use relatively short sentences in conversations. As a result, the difficulty of narrative text is typically underestimated. Such underestimations are most evident in texts such as the classic *Old Man and the Sea* (Hemingway, 1952) receiving a Lexile of 940 (which falls into the grade 4–5 band of the recalibrated Lexile Levels).

A third criticism has to do with the reliance on a single score for a text. On the Lexile Map (MetaMetrics, 2000), *Pride and Prejudice* (Austen, 1813) is given as a prototype for 1100 Lexile, and *Modern Biology* (Holt, Rinehart & Winston, 1999) for 1130 Lexile. These two texts are judged to be relatively the same in comprehensibility. However, the variability across individual parts of texts can be extensive. Within a single chapter of *Pride and Prejudice*, for example, 125-word excerpts of text had Lexiles that ranged from 670 (beginning grade three) to 1310 (college).

There is at least one criticism that is unique to Lexiles, reflecting the manner in which the semantic component is established. The semantic measure comes from the average log frequency of words in a text. The distribution of words in English (or any language) is extremely skewed. Approximately 1,000 words account for less than 1% of the words in written English but this group of words accounts for approximately 67% of the total words in text. At the other end of the distribution, approximately 60% of the words in English appear less than once per million words of text (Leech, Ray, & Wilson, 2001). When so many

words have the same rating, the predictive validity of the Lexiles can be limited.

When a document endorsed by the majority of the nation's states gives specific readability levels (as is the case with the CCSS/ELA), the opportunity for misinterpretations of readability data are many. Chall, one of the developers of the longest and most widely used readability formula (Chall & Dale, 1995; Dale & Chall, 1948) consistently noted that there is nothing inherent in the formula itself that leads to misuse (Chall, 1985; Chall & Dale, 1995). Chall's admonition serves as an impetus for the current analysis—identifying and applying a set of quantitative measures, rather than a single measure to texts that have been offered as appropriate for particular grade bands.

Identifying a Set of Quantitative Measures of Text Complexity

The possible quantitative measures that have been proposed for the analysis of text difficulty are many. The Coh-Metrix framework (Graesser, McNamara, Louwerse, & Cai, 2004), for example, provides data on 62 quantitative measures of text cohesion and text difficulty, although the unique contributions of each measure to reading proficiency have yet to be untangled.

The aim of the analyses I conducted is to contrast the Lexile score of a text with data gained from three measures. Two of these measures are the constituents of a Lexile score that have already been described: MSL and MLWF. The third measure is a central one in the Coh-Metrix framework—referential cohesion.

Intra-Lexile Measures: MSL and MLWF

Typically, Lexiles are reported as an overall figure, ranging from 0 to 2,000 but Information on the two constituents that form the Lexile—the MSL and MLWF—are part of the output of an analysis at www.lexile.com. In this paper, MSL and MLWF are referred to as “intra-Lexile” measures.

Intra-Lexile data for the two texts that were the source for the excerpts at the beginning of this paper illustrate how the measures work. The figures for MSL are: 9.2 words for *Art Around the World (Art)* and 8.4 words for *Sarah: Plain and Tall (Sarah)*. The means indicate that Art has slightly longer sentences than Sarah.

The MLWF data are 3.35 for *Art* and 3.84 for *Sarah*. The MLWF is relatively easy to interpret in a comparison such as this. Given that a lower number means that the words of a text, overall, are less frequent, these means indicate that *Art* has less frequent vocabulary than *Sarah*. But the MLWF could be low or high for a number of reasons. For example, a single word that is very rare, such as *Mudge*, might be repeated up to 30 times in a 750-word text, as is the case in *Henry and Mudge* (Rylant, 1987). However, few other words in that text are rare. A low MLWF also may be the result of numerous, rare words in a text, all of which appear a single time. This is the case with *The Birchbark House*

TABLE 2
Typical Ranges for Word Frequency and Sentence Length

Grade Band	Narrative Texts		Informational Texts	
	Word Frequency	Sentence Length	Word Frequency	Sentence Length
2	3.7-3.9	8-10	3.6-3.8	9-11
3	3.6-3.8	9-11	3.5-3.75	10-12
4	3.5-3.8	10-12	3.4-3.6	11-13
5	3.4-3.7	11-13	3.3-3.6	12-14

(Erdich, 1999). These two patterns of rare vocabulary place quite different demands on students' vocabulary prowess.

Another issue has to do with the range of this measure. Theoretically, the measure can range from 0 to 5, but the typical range in children's texts from grades two through five is limited to 3.0 to 3.9. The reason for this limited range has already been described: the presence of thousands of words that appear very infrequently in written English.

When comparing one text against another, a conclusion of "harder" or "easier" may be sufficient. Yet, as educators select appropriate texts, they want to know what typical ranges of these measures are for particular grade levels. Such guidelines have not yet been provided at www.lexile.com. To support interpretation of the factors that may be teachable in a text, particularly vocabulary, data on the range of MSL and word frequency are provided in Table 2. Readers are cautioned to interpret these data as preliminary in that they are based on an analysis of only 200 texts (50 at each grade level). Further, these guidelines do not answer questions about what "long" sentences" or "low" word frequency means for students' comprehension and instruction. They are provided to give at least a preliminary framework for interpreting quantitative data on text complexity.

Referential cohesion

The CCSS/ELA description of quantitative indices of text identifies text cohesion as a critical component of text complexity, but does not include recommendations as to how these data could be gathered. Halliday and Hasan (1976) identified two main types of cohesion: grammatical, which refers to the structural content, and lexical, which refers to the language content. Numerous sub-types exist within each group, such as referential cohesion within the lexical group. Referential cohesion has proven particularly predictive of the demands on elementary students' comprehension (McNamara, Graesser, Cai, & Kulikowich, 2011).

Referential cohesion refers to overlap in content words between sentences within paragraphs or sections of a text. One way in which this overlap can be measured is to determine whether nouns, pronouns, and noun-phrases are re-

peated across sentences of a text. The two excerpts that follow illustrate different degrees of this form of cohesiveness.

Example A: A seed is where most plants begin life. There are other ways plants can begin life, but most plants begin as seeds. (from *Seed to Plant*, Gibbons, 1991)

Example B: A black nose sniffs the air. Then a smooth white head appears. A mother polar bear heaves herself out of her den. (from *Where Do Polar Bears Live?* Thompson, 2010)

The high level of cohesiveness in the first text, as indicated by a score of .86 (on a scale of 1 to 0) is apparent in the reference to “seeds” in adjacent sentences. The score of .11 for cohesiveness in the second example underscores the inference that is required of young readers (i.e., that the nose and head belong to the mother polar bear).

The two following excerpts illustrate a second type of content overlap—stem overlap which refers to the degree to which words with the same root or stem appear in a text.

Example C: “Not a short one,” he said. “Not a curly one,” he said. And no pointy ears. Then he found Mudge. Mudge had floppy ears, not pointed. (from *Henry & Mudge: The First Book*, Ryland, 1987)

Example D: “Ordinarily I’d save you for afternoon tea, but I happen to be upset enough and hungry enough to eat you right now.” And he picked up my father in his front paws to feel how fat he was. (from *My Father’s Dragon*, Gannett, 1948)

Example C illustrates a moderately high level of stem overlap (.70 on a scale of 1 to 0) as illustrated by the presence of *pointy* and *pointed*, while no derivatives of the same word are shared in Example D, leading to a low level of stem overlap (.05 for the entire text).

Numerous questions remain about the precise effects of cohesion on comprehension (Deane, Sheehan, Sabatini, Futagi, & Kostin, 2006). For example, struggling readers have been found to rely on strong cohesion to a greater degree than proficient readers (McNamara & Kintsch, 1996). Further, cohesiveness likely has different characteristics in informational than in narrative texts. Initial analyses of referential cohesion, however, are sufficiently promising to include it in a quantitative analysis of text complexity (Hiebert & Pearson, 2010).

Referential cohesion in this analysis combines two Coh-Metrix measures—noun/pronoun/noun-phrases and shared root words (Graesser et al., 2004). Researchers on the Coh-Metrix project are developing a Coh-Metrix Easability Components program (McNamara et al., 2011) is intended to provide guidance on acceptable levels of referential cohesion at different levels and for different

TABLE 3
Typical Averages for Referential Cohesion (Argument and Stem)

	Narrative Texts	Informational Texts	Grade 2–3 Band	Grade 4–5 Band
Argument Overlap	.32	.56	.40	.46
Stem Overlap	.23	.55	.37	.38

genres. Until that program is available, the summary in Table 3 gives insight into typical levels of texts that have been deemed exemplary (i.e., all of the exemplars identified within Appendix B of the CCSS/ELA for grades 2 through 5). The data, summarized in Table 3, show that referential cohesion differed substantially by genre. The levels of cohesion are substantially higher in informational than in narrative texts. Within the same genre, the levels of referential cohesion vary little between the two grade bands. When the referential cohesion of a text is substantially discrepant from these levels, as was the case with *My Father’s Dragon* (a level of .05), such information is critical to consider relative to other sources of data.

Applying the Quantitative Measures to Sets of Texts

Six exemplars from the CCSS/ELA list for the grade 2–3 band, three narrative and three informational, were chosen from the lower end of the designated range for that grade band: 430–680. Although this range may appear considerable (in that 100 Lexiles are equivalent to a grade level), the three informational texts had the lowest Lexiles of the informational texts in the CCSS/ELA pool for this grade band. For the CCSS/ELA exemplars for grade 4–5, three narrative and three informational texts were in the 820–890 Lexile range (i.e., approximately a half-grade level to one another). The patterns on the grade 2–3 texts are reported in Table 4 and those for the grade 4–5 texts in Table 5.

Grade 2–3 exemplar texts

The data for variables have been ranked to allow for comparison across a set of texts. Several observations can be made about the data in Table 4. First, the use of a single measure, whatever the measure, leads to quite different interpretations of text complexity. If the single criterion is the Lexile, *Sarah* (MacLachlan, 1985) would be assigned to the below-basic readers in a second-grade class and *Fire Cat* (Averill, 1960) would be assigned to more proficient readers. If the single criterion is referential cohesion, *Sarah* would be viewed as appropriate for proficient readers at the end of second grade, if not third grade, and *Fire Cat* would be viewed as appropriate for below-basic second-grade readers.

Second, several variables appear to be interchangeable, while others appear to provide unique information. Data in Table 6 confirm the close alignment of Lexiles and MSL variables as evident in the correlation of .97 between these two variables. Hiebert (2010) reported a correlation for Lexile and MSL in a

TABLE 4
Quantitative Indices: Exemplar Texts from CCSS Grade Band 2–3

Genre	Title	Lexile		Sentence Length		Word Frequency		Referential Cohesion	
		Score	Rank ¹	Score	Rank ¹	Score	Rank ¹	Score	Rank ¹
Informational	Art Around the world	680	6	9.2	6	3.35	6	.39	4
	Bats—Creatures of the Night	450	2	7.5	1	3.55	5	.60	2
	Martin Luther King	560	5	9.1	5	3.65	3.5	.30	5
Narrative	Fire Cat	480	4	8.7	4	3.76	2	.54	3
	Henry & Mudge	460	3	8.0	2	3.65	3.5	.71	1
	Sarah: Plain & Tall	430	1	8.4	3	3.84	1	.23	6

¹ 1=easiest; 6=hardest

similar range for the entire sample of second through fifth-grade texts on the CCSS exemplary list: .86.

The correlation of the MLWF to the Lexile, $-.76$, is higher than the $-.51$ found in the entire sample of CCSS second through fifth grade texts (Hiebert, 2010) but it is not at the level of the MSL and Lexile relationship. The referential cohesion measure does not have a strong relationship to any of the other measures, suggesting that it may contribute unique information to understanding text complexity (or it may have insufficient reliability).

Finally, the variability in text characteristics is considerable between narrative and informational texts. This variability may be an artifact of the sample—both of this study and of the CCSS/ELA. With respect to this study, the easiest texts within the grade 2–3 band were the focus. The degree of variability may be less in the higher ranges of text represented in this grade band. With respect to the sampling of the CCSS/ELA, the recalibration of the Lexiles resulted in increasing the range of text to be covered during grades two and three. Over this two-grade span, the Lexile range of 450 to 790 represents almost 3.5 grade levels (100 Lexiles are described as a grade level). Books have been offered to teachers and policy-makers as exemplars with no differentiation for this enormous grade range. Teachers are advised that, with their scaffolding, second graders should be able to read even the most complex texts offered for this band. But the changes in reading over this period are more than quantitative in nature—as measured by the ability to read longer sentences or harder and more words. Chall’s (1983) stages describe massive changes over this developmental period where children move from, initially, attending to the code, then becoming automatic with a vast vocabulary, and, at the end of this period, using their reading acumen to acquire information from text.

Yet another feature of the CCSS/ELA sampling procedures needs to be considered in considering the variability: their decision to exclude any books published by the school divisions of publishing houses. Numerous informational

TABLE 5
Quantitative Indices: Exemplar Texts from CCSS Grade Band 4–5

Genre	Title	Lexile		Sentence Length		Word Frequency		Referential Cohesion	
		Score	Rank ¹	Score	Rank ¹	Score	Rank ¹	Score	Rank ¹
Informational	Kenya	820	1	11.6	1	3.43	3	.57	2
	History of US	880	4.5	12.5	4	3.42	4	.35	5
	Hurricanes	880	4.5	13.2	5	3.53	1	.68	1
Narrative	The Birchbark House	860	2.5	11.9	2	3.39	6	.38	4
	Tuck Everlasting	860	2.5	12.0	3	3.40	5	.20	6
	M.C. Higgins the Great	890	6	13.3	6	3.51	2	.48	3

¹ 1=easiest; 6=hardest

books that are appropriate for beginning readers exist—ones that are considerably easier than the informational texts in the CCSS/ELA exemplar list (Duke & Bennett-Armistead, 2003).

Grade band 4-5

In contrast to the grade 2–3 sample of texts, it was possible to get a set of narrative and informational texts from the CCSS/ELA exemplar list for grades 4–5 that fell into a limited Lexile range, as is evident in Table 5. Even among texts that have a Lexile range within about one-half grade level, the application of alternative quantitative indices shows substantial differences in text complexity. This is especially true for the referential cohesion measure. The text that is evaluated as most accessible according to the referential cohesion measure is *Hurricanes* (Lauer, 1996)—the text that has the highest Lexile of the group. *Hurricanes* is written in a fairly straightforward manner. Its MLWF also indicates that it has the most accessible vocabulary. On the other hand, *Tuck Everlasting* (Babbitt, 1975) and *The Birchbark House* (Erdich, 1999) both have Lexiles in the lower range but have MLWFs that indicate the presence of challenging vocabulary. Further, the referential cohesion indices are low, suggesting that students will need to make numerous inferences. To an even greater degree than the exemplar texts in the grade 2–3 band, the assignments of Lexiles to texts in the grade 4–5 band confirm the observation that readability formulas tend to underestimate the difficulty of narratives and overestimate the difficulty of informational texts.

TABLE 6
Relationships among Measures

	Lexile	MSL	MLWF
MSL	0.97		
MLWF	-.076	-.58	
RefCoh	-0.12	-0.09	0.07

Using Our Professional Expertise and Experience

As I noted earlier, quantitative information about a text is useful as a way to get a general sense of a text's difficulty, especially when choosing among many texts. However, as my analyses show, the use of only one quantitative measure, such as Lexiles, can produce unintended consequences. Just a quick review of book covers can raise questions about text assignments based solely on Lexiles. For example, the Newbery Award seal on the cover of *Sarah* (MacLachlan, 1985) signals that the story is likely a sophisticated one that will require readers to make numerous inferences. The requisite level of inferencing (confirmed by the referential cohesion rating) would make this text challenging for below-basic second graders (as suggested by the Lexile). By contrast, the "An I Can Read" designation on the cover of *Fire Cat* (Averill, 1960) would encourage teachers to examine this book more closely. A somewhat higher Lexile notwithstanding, teachers could quickly identify the words that require instruction for students to be successful with texts. Moreover, even among a set of quantitative measures, outcomes describing the potential difficulty of a text can vary considerably. As the analyses show, even among texts that have a Lexile range within about one-half grade level, the application of alternative quantitative indices shows substantial differences in text complexity.

Readability formulas and quantitative data have a place in the evaluation of text complexity. As Chall (1985; Chall & Dale, 1995) observed, the problems with readability formulas lie with interpretation and use, not with the formulas themselves. Quantitative data from readability formulas requires the same review and thought that we might give to addressing a child's high temperature. Before we make dramatic decisions about treatment based on that temperature, we should apply additional forms of measurement (as well as recognize that factors such as time of day and location influence temperature readings). And we need to understand that, even with multiple temperature readings, these data do not indicate *what* the child's problem is. Before choosing to treat the child with chemotherapy (cancerous substances can cause high fevers), we first undertake numerous tests and consider alternative causes for the fever—the child might have an infection or may be reacting to a medication. Each of these causes calls for a different treatment.

Similarly, *quantitative* data requires verification and inclusion of *qualitative* forms of data, including information about the needs and strengths of the students who are reading the texts and the support structures and the tasks around the reading of the texts.

What I hope educators will take away from these analyses is a clear understanding that quantitative measures such as Lexiles are a good place to start as they make decisions about matching texts to students' reading abilities. But once they have data from these measures, they must elaborate their findings with qualitative information about individual students and books.

To help educators clarify issues to consider in determining the best-student text matches and so support students' reading development, educators clearly need additional tools and procedures. Projects currently underway may provide this help. For example, the *Coh-Matrix Easability Components* (McNamara et al., 2011) system is a promising source of additional quantitative data, and additional qualitative guidance may emerge from several projects currently underway (Liben & Liben, 2011).

The Common Core State Standards offer a positive step toward urgently needed reform in our schools. The standards' focus on the issue of text complexity is an important part of that reform, leading us to look closer at the texts we expect our students to read and at the support we give them to learn from those texts. Matching students to appropriate texts is a crucial part of this support. To this end, we must continue to develop and make available for educators the tools and procedures they need to make the best possible text decisions for their students.

References

- 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: The Center for the Study of Reading, National Institute of Education, National Academy of Education.
- Beck, I.L., McKeown, M.G., Omanson, R.C., & Pople, M.T. (1984). Improving the Comprehensibility of Stories: The Effects of Revisions That Improve Coherence. *Reading Research Quarterly*, 19(3), 263–277.
- California English/Language Arts Committee. (1987). *English-Language arts framework for California public schools (kindergarten through grade twelve)*. Sacramento, CA: California Department of Education.
- Carroll, J.B., Davies, P., & Richman, B. (1971). *The American heritage word frequency book*. Boston: Houghton Mifflin.
- Chall, J.S., & Dale, E. (1995). *Readability revisited: The new Dale-Chall readability formula*. Cambridge, MA: Brookline Books.
- Chall, J. S. (1985). Afterword. In R. C. Anderson, E. H. Hiebert, J. A. Scott, & I. A. G. Wilkinson (Eds.), *Becoming a nation of readers* (pp. 123–125). Champaign, IL: The Center for the Study of Reading, National Institute of Education, National Academy of Education.
- Chall, J.S. (1983). *Stages of reading development*. New York, NY: McGraw-Hill Book Company.
- Cohen, S.A., & Steinberg, J.E. (1983). Effects of three types of vocabulary on readability of intermediate grade science textbooks: An application of Finn's transfer feature theory. *Reading Research Quarterly*, 19(1), 86–101.
- 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.
- Dale, E., & Chall, J.S. (1948). A formula for predicting readability and instructions. *Educational Research Bulletin*, 27, January 21 & February 18, 1948, pp. 11–20, 28, & 37–54.
- Deane, P., Sheehan, K.M., Sabatini, J., Futagi, Y., & Kostin, I. (2006) Differences in Text Structure and Its Implications for Assessment of Struggling Readers, *Scientific Studies of Reading*, 10(3), 257–275.
- Duke, N.K., & Bennett-Armistead, V.S. (2003). *Reading and writing informational text in the primary grades*. New York, NY: Scholastic.
- Finn, P.J. (1978). Word frequency, information theory, and cloze performance: A transfer theory of processing in reading. *Reading Research Quarterly*, 13, 508–537.
- Graesser, A. C., McNamara, D. S., Louwerse, M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavioral Research Methods, Instruments, and Computers*, 36, 193–202.
- Gray, W.S., & Leary, B. (1935). *What makes a book readable*. Chicago: University of Chicago Press.
- Green, G., & Davison, A. (Eds.). (1988). *Linguistic complexity and text comprehension: Readability issues reconsidered*. Hillsdale, NJ: Erlbaum.
- Halliday, M.A.K., & Hasan, R. (1976): *Cohesion in English*. London: Longman.
- Hiebert, E.H. (December 8, 2010). *The view of text complexity within the Common Core State Standards: What does it mean for struggling readers?* Plenary address at the American Research Forum, Sanibel Island, FL.
- Hiebert, E.H., & Pearson, P.D. (2010). *An examination of current text difficulty indices with early reading texts* (Reading Research Report 10.01). Santa Cruz, CA: TextProject, Inc.
- 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.
- Leech, G., Rayson, P., & Wilson, A. (2001). *Word frequencies in written and spoken English based on The British National Corpus*. London: Longman.
- Liben, D., & Liben, M. (April 10, 2011). *The emergence of the active ingredients of text: A unique marriage of a quantitative and qualitative research effort*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.
- Lively, B. A., & Pressey, S. L. (1923). A method for measuring the vocabulary burden of textbooks. *Educational Administration and Supervision*, 9, 389–398.
- McNamara, D.S., & Kintsch, W. (1996). Learning from texts: Effects of prior knowledge and text coherence. *Discourse Processes*, 22, 247–288.
- McNamara, D.S., Graesser, A.C., Cai, Z., & Kulikowich, J.M. (April 9, 2011). *Coh-Metrix Easability Components: Aligning Text Difficulty with Theories of Text Comprehension*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.
- RAND Reading Study Group (2002). *Reading for understanding: Toward an R & D program in reading comprehension*. Santa Monica, CA: RAND Science and Technology Policy Institute.

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. ED 307 577)

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

Literature

Austen, J. (1813). *Pride and Prejudice*. [Retrieved on October 20, 2008, from <http://www.authorama.com/pride-and-prejudice>]

Averill, E. (1960). *The Fire Cat*. New York, NY: HarperCollins.

Babbitt, N. (1975). *Tuck Everlasting*. New York, NY: Farrar, Straus and Giroux.

Erdrich, L. (1999). *The Birchbark House*. New York, NY: Hyperion.

Gannett, R. S. (1948). *My Father's Dragon*. New York, NY: Random House.

Gibbons, G. (1991). *From Seed to Plant*. New York, NY: Holiday House.

Hakim, J. (2005) *A History of US*. Oxford, UK: Oxford University Press, 2005.

Hamilton, Virginia. M. C. *Higgins, the Great*. New York, NY: Simon & Schuster, 1999.

Hemingway, E. (1952). *Old Man and the Sea*. New York, NY: Scribner.

Holt, Rinehart & Winston (1999). *Modern Biology*. New York, NY: Random House Value Publishing.

Lauber, P. (1996). *Hurricanes: Earth's Mightiest Storms*. New York, NY: Scholastic.

Leonard, H. (1998). *Art Around the World*. New York, NY: Rigby.

MacLachlan, P. (1985). *Sarah, Plain and Tall*. New York, NY: HarperCollins.

Milton, J. (1993). *Bats: Creatures of the Night*. New York, NY: Grosset & Dunlap.

Ruffin, F. E. (2000). *Martin Luther King and the March on Washington*. New York, NY: Grosset & Dunlap.

Rylant, C. (1987). *Henry and Mudge: The First Book of Their Adventures*. New York, NY: Atheneum.

Thomson, S. L. (2010). *Where Do Polar Bears Live?* New York, NY: HarperCollins.

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

Editorial Board for TextProject, Inc. Publications

Martha Adler

University of Michigan, Dearborn

Victoria Appatova

University of Cincinnati

Kathie Bach

Apex Learning

Suzanne Barchers

Stanford, CA

Alison Billman

*Lawrence Hall of Science/
University of California, Berkeley*

Marco Bravo

Santa Clara University

Devon Brenner

Mississippi State University

Janelle Cherrington

Scholastic

Janet Gaffney

*University of Illinois,
Champaign-Urbana*

Robert Gaskins

Benchmark School

Shannon Henderson

University of Arkansas, Little Rock

Heather Koons

Metametrics

Melanie Kuhn

Boston University

Pamela Mason

Harvard University

Shailaja Menon

Jones International University

Heidi Anne Mesmer

Virginia Tech

Maria Murray

SUNY-Oswego

Colleen Klein Reutebuch

University of Texas, Austin

Paula Schwanenflugel

University of Georgia

Alexandra Spichtig

Reading Plus

Guy Trainin

University of Nebraska, Lincoln

Masa Uzicanin

Wireless Generation

Claire White

Harvard University

Kathy Wilson

University of Nebraska, Lincoln