



---

## **Interpreting Lexiles** in Online Contexts and with Informational Texts

**Elfrieda H. Hiebert**  
University of California, Berkeley

February 2009

<b>Contents</b>	
<b>Introduction .....</b>	<b>3</b>
<b>Readability Formulas .....</b>	<b>5</b>
Descriptions .....	5
Critiques .....	5
<b>Lexiles .....</b>	<b>6</b>
Description .....	6
Critiques of Lexiles .....	7
<b>Recommendations .....</b>	<b>9</b>
<b>References .....</b>	<b>11</b>

## Introduction

**Excerpt 1:** Today, chemists use their understanding of the behavior of atoms to imagine and then create new compounds that have never existed in nature. Since Fleming's day, chemists have designed many different antibiotics to overcome the growing problem of antibiotic resistance. Molecular computer modeling allows scientists to create designer chemicals that serve specific purposes in medicine, manufacturing, or even art. (Chemistry course, Apex Learning, 2008)

**Excerpt 2:** "He is, indeed; but, considering the inducement, my dear Miss Eliza, we cannot wonder at his complaisance, for who would object to such a partner?" Elizabeth looked archly, and turned away. Her resistance had not injured her with the gentleman, and he was thinking of her with some complacency, when thus accosted by Miss Bingley, "I can guess the subject of your reverie." (Jane Austen, *Pride and Prejudice*, 1813)

How different are these texts in their comprehensibility for readers? This question is a critical one—not only for schools but also for businesses in manuals given to customers, for the military in directions that are given to personnel, or for federal agencies such as the Internal Revenue Service in forms that citizens need to file.

For almost a century, readability formulas have been the primary way of describing the differences in texts such as these (Klare, 1984). According to a "classic" readability formula—the Flesch-Kincaid (Flesch, 1948; Kincaid, Fishburne, Rogers, & Chissom, 1975)—Excerpt 1 is at Grade 12 and Excerpt 2 is at Grade 8.9. In the 1980s, research findings challenged underlying assumptions of classic readability formulas, such as the idea that short sentences are always easier (Anderson, Hiebert, Scott, & Wilkinson, 1985). For example, by changing the punctuation in Excerpt 2, the number of words per sentence drops from 22 to 8, and the readability drops by three grade levels—from 8.9 to 5.9. It is doubtful, however, whether these changes would make it easier for a sixth grader (or even a ninth grader) to understand the haughtiness of Miss Bingley or the shift in Mr. Darcy's affections. While use of classic readability formulas in the creation of core reading programs has declined over the past 25 years, a "new" generation of readability formulas has become prominent. These new readability formulas use digital technology that permits ease both in providing data on thousands of texts and in developing text-specific, online assessments.

Foremost among this new generation of readability formulas is the Lexile Scale (LS) (Smith, Stenner, Horabin, & Smith, 1989). The LS has become so ubiquitous as an index for texts that many states and school districts expect it to be reported on instructional materials, including texts and digital products. Texts in digital contexts have unique features that are not present in typical classroom contexts. Furthermore, when the texts are informational, as they often are in online contexts, the need to adjust readability formulas has long been recognized (Cohen & Steinberg, 1983; Finn, 1978).

In this paper, I present a brief overview of the history and content of readability formulas, including critiques that apply to the LS as well as to the classic readability formulas. Next, I describe the LS and criticisms specific to it. The final section of the paper considers ways in which the publishers can respond to the ubiquitous demand for correlation of texts and materials to the LS. But first, I provide two needed caveats.

The first caveat pertains to the first phrase of the title, “in online contexts.” The research literature on the use of either classic or new readability formulas with texts in online contexts is sparse, almost nonexistent. In one study, Ta-Min (2006) applied a variety of readability formulas (but not the LS) to a Web site on cancer prevention. Ta-Min concluded that results of the readability formulas were unreliable in evaluating the Web content and were not associated with measures of the coherence of Web pages. Ta-Min concluded that the evaluation of content in digital contexts requires measures of comprehensibility much more sophisticated than those of the classic or new readability formulas. The factors that influence comprehension of text in digital settings are sufficiently complex and numerous to make the topic the source of a subsequent paper (Hiebert, Menon, & Martin, in progress).

The second caveat pertains to the second phrase of the title, “with informational texts.” Both the classic and new readability formulas have been developed and tested almost exclusively with narrative texts. For example, there were few, if any, words within the Dale and Chall (1948) or Spache (1974) lists that represented the content of science or social studies. Soon after readability formulas began to be used extensively, Burkey (1954) identified the need to adapt the formulas for informational texts, where, unlike narrative texts, key words are repeated. As Biber (1988) has noted, the differences across the genres of “story” and “information” are as substantial as—if not more so than—the differences across oral and written language. Narrative genres often use numerous common words between the rare words, as occurs in Excerpt 2 (*looked*, *turned*, and *guess* for the former and *inducement*, *complaisance*, and *reverie* for the latter). Common words in any language are typically short—a factor in readability formulas (Zipf, 1935). Informational texts have common words (e.g., articles, prepositions), but these texts also use a general academic vocabulary (GAV) (Coxhead, 2000) to provide precision and specificity. Words such as *behavior* and *existed* in Excerpt 1 are illustrative of GAV vocabulary—multisyllabic words that occur with less frequency than common words such as *looked* and *guess*. Further, dialogue in narrative texts is often represented by short sentences, while the sentences of informational genres contain phrases that explicate points. As with their predecessors, the new readability formulas have failed to address how the differences in text features and structures of genres, as well as the dispositions and knowledge of readers, can influence text difficulty. One of the aims of this paper is to raise issues in interpreting the difficulty of informational text.

## Readability Formulas

### Descriptions

Lively and Pressey's (1923) formula is often cited as the first readability formula, but it was Gray and Leary's analysis (1935) that laid the foundation for the type of readability formulas that became prominent. Gray and Leary clustered the 228 factors in text difficulty that librarians, teachers, and professors identified into four groups: content, style, format, and features of organization. After giving texts from magazines, newspapers, and books to a sample of 800 adults, Gray and Leary concluded that content, format, and organization could not be evaluated statistically. Of the 80 variables related to style, Gray and Leary identified 64 that could be reliably counted. Of the 64 countable variables, none had a higher correlation than 0.52 in explaining readers' performances. To achieve a better statistical result, Gray and Leary combined five variables that had a correlation of 0.645 with reading difficulty: average sentence length, number of different hard words, number of first-, second-, and third-person pronouns, percentage of different words, and number of prepositional phrases.

Other researchers established that the first two variables—a semantic (meaning) measure, such as difficulty of vocabulary, and a syntactic (sentence structure) measure, such as average sentence length—were the best predictors of textual difficulty. Average number of words per sentence became the typical syntactic index, while the semantic component was measured in one of two ways. One is typified in Dale and Chall's (1948) formula, in which words were regarded as semantically complex if they were not present on a specified list of useful words. The second is exemplified by Flesch's (1948) formula, in which semantic complexity was based on the number of syllables and/or letters. By the early 1980s, Klare (1984) identified around 200 readability formulas and well over a thousand studies on text readability.

### Critiques

Criticisms of readability formulas came soon after their development (Moore, 1935). The criticisms fall into three groups, the first concerning the validity of the two constructs—word difficulty and sentence length—that represent text difficulty. From the vantage point of cognitive processing, short sentences can mean that conjunctions and phrases that make a text more coherent are lacking (Green & Davison, 1988). Further, texts that have common words (e.g., *drugs* rather than *antibiotics* in Excerpt 1) can mean that readers' opportunities for meaningful learning are limited (Green & Davison, 1988).

A second criticism of readability formulas was identified in the introduction of this paper—their applicability to content-area texts, especially science texts. While studies have not been conducted on the validity of the new readability formulas with science texts, classic readability formulas have produced higher grade-level assignments than the grades assigned by publishers (Cramer & Dorsey, 1969). Walker (1965) showed that, while science texts contained many rare words, these rare words were repeated more frequently than was the case in narrative texts. Finn (1978) found that words below particular frequency levels tended to be repeated more frequently in science texts than were words in the middle range. Readers were able to supply these repeated words in word-deletion cloze tests of the science texts more easily than unrepeated rare words. Finn reasoned that repetition causes these words to receive extraordinary transfer-features support from surrounding words. Cohen and Steinberg's (1983) analyses of science textbooks confirmed

Finn's suppositions: the failure of readability formulas to identify technical, unfamiliar words that were repeated, and to treat those words differently from other unfamiliar words, resulted in unreliable indicators of the readability of science texts.

The third criticism pertains to the uses that readability formulas came to have. The concentration over decades on one approach to establish readability and the ease of this approach meant that, by the 1980s, there was a substantial amount of abuse. When state departments of education or large city school districts mandated particular readability requirements, publishers began to manipulate texts to fit the requirements of formulas, including creating new texts that used shorter sentences and/or common rather than rare, multisyllabic words (Davison & Kantor, 1982). By the early 1980s, there was recognition that readability formulas had influenced school texts, not only those used to teach young children to read, but also those used to teach secondary students content in biology, chemistry, and physics.

Critiques of the role of readability formulas in creating or manipulating texts were communicated to practitioners through documents such as *Becoming a Nation of Readers* (Anderson et al., 1985) and *Learning to Read in American Schools* (Anderson, Osborn, & Tierney, 1984). On the basis of these critiques, influential states that adopted reading programs barred the use of readability formulas in establishing selections in basal reading programs (California English/Language Arts Committee, 1987; Texas Education Agency, 1990), influencing the text difficulty of reading programs nationwide.

---

## Lexiles

---

### *Description*

At the same time that reading researchers were describing the limitations of readability formulas, several efforts were under way in which readability formulas were being delivered digitally: Advantage-TASA Open Standard Readability Formula for Books (ATOS) (School Renaissance Institute, 2000) and the Lexile Scale (LS) (Smith et al., 1989). Like their predecessors, the new readability formulas use some combination of syntactic and semantic measures to attain readability. However, the new readability formulas differ in implementation from classic readability formulas in several important ways. Most significantly, the use of digital technology meant that the readability of many texts could be established—something that had been tedious and difficult to accomplish with the time-intensive earlier procedures. Further, developers retained the processing of readability as intellectual property, requiring educators and other clients to pay for their services to obtain readability levels. The digital technology also meant that developers could offer sections of texts as online assessments. As a result, the new readability formulas acquired an application that had not been claimed for classic readability formulas: the application of the same metric could be used to establish students' reading levels, and the right match to texts could be recommended through these digital analyses. Designers of classic readability formulas such as the Dale-Chall, even in the revised version of the formula (Chall & Dale, 1995), made no claim that the formulas could be used to establish the proficiency of the reader.

The ATOS (School Renaissance Institute, 2000) has focused on sales of an independent reading program where students' selections are proscribed by a readability formula. MetaMetrics, the company that owns the LS, has focused on selling its services to publishers

of tests and texts. As a result of this focus, all major publishers of school and trade texts for the educational market provide Lexiles on their materials, and the LS is used on national assessments (MetaMetrics, 2003). Further, numerous states require information on the LS for purchase of materials with state funds. While the Flesch-Kincaid readability formula, which is part of Microsoft Word software, is used universally for English texts, the major method of establishing text difficulty in K-12 education in American schools is the LS.

According to Stenner, Burdick, Sanford, and Burdick (2006), text difficulty according to the LS is established by theory rather than empiricism. The theory component appears to be the use of Carroll, Richman, and Davies's (1971) word-frequency data for the semantic component. The Carroll et al. corpus consists of the relative frequencies of 5 million words from 1,045 published titles that were commonly used in grades three through nine in the 1960s.

A word's frequency (represented as mean log word frequency) is analyzed with a measure of the number of words per sentence to produce a score on a scale from 0 to 2000, spanning beginning reading through the most complex forms of reading. According to The Lexile Framework for Reading (MetaMetrics, 2000), the levels that span high school are as follows: (a) 9th grade: 1050-1150, (b) 10th grade: 1100-1200, (c) 11th grade: 1120-1210, and (d) 12th grade: 1210-1300. To demonstrate the data provided by the Lexile Analyzer, the two excerpts that introduced this paper have LSs of 1260 (Excerpt 1) and 920 (Excerpt 2). Thus, the first excerpt falls into the 12th grade, and the second excerpt falls within the ranges of sixth to eighth grades.

A reader is assigned a level based on correctly answering 75% of the cloze questions, which are based on the 125-word calibrations in a text. A reader with a measure of 1100L who is given a text measured at 1100L is expected to have a 75% comprehension rate. The test uses a cloze-like technique in which every *n*th word is deleted and students are presented with the target word and three additional words that are syntactically correct, and they are asked to choose the word that best completes the sentence, given the meaning of the text.

### ***Critiques of Lexiles***

The criticisms that are applied to classic readability formulas are also applicable to the LS, such as the reliance on simplistic indices of text difficulty and the lack of adjustments for technical vocabulary. The LS is even more vulnerable than classic readability formulas are to criticisms related to the use and abuse of the information, due to claims by its developers that readers can be matched precisely with texts (Smith et al., 1989).

There is at least one criticism that is unique to the LS because of the form of the semantic measure. Marilyn Adams, a member of the panel that the National Center for Education Statistics (NCES) convened to determine the usability of the LS for national assessments (White & Clement, 2001), raised a potential problem with the discriminating power of Lexiles in the middle ranges. Adams argued that, because of the large number of words with frequencies in the range of approximately one occurrence per million words of text (approximately 60% of the words within the British National Corpus), many passages might have equivalent Lexile values (see also Adams, in press). When so many words have the same rating, the discrimination of the LS may be limited. The panel recommended that MetaMetrics experiment with some other way (either mathematical or semantic) of defining word frequency on the LS that would provide more discrimination in the middle ranges.

There is no indication that MetaMetrics has engaged in such experimentation. A preliminary study by Hiebert (2008), however, confirms the findings of the NCES Panel report regarding the indiscriminability of the LS with vocabulary in the middle ranges (White & Clement, 2001). Hiebert modified an excerpt of *Pride and Prejudice*, which is offered as the prototypical text for 1100 Lexile (MetaMetrics, 2000)—a point on the LS that is within the range of most high-school grades (9-11). Hiebert substituted words in the original text that had predicted appearances of less than once per million to form three alternative texts: (a) Semantic #1, in which the substitutions consisted of words with predicted appearances of 1 to 9 occurrences per million, (b) Semantic #2, with substitutions of words with predicted occurrences of 10 to 99 times per million, and (c) Semantic #3, with predicted occurrences of 100 to approximately 10,000 times per million. The difference between the original text and Semantic #3 was 50 Lexiles—approximately 30% of a grade level. By contrast, the difference on a classic readability formula (the Flesch-Kincaid) from the original to Semantic #3 was 1.2 grade levels. Even though words such as *daydream* and *disapproval* are likely to be much more familiar to a ninth grader than *reverie* or *strictures* (see Dale & O'Rourke, 1981), the first pair of words is relatively rare in written English. According to the LS, the differences between these two pairs of words are not predicted to be great. Adams's observations regarding the lack of discriminability of the word-frequency indices for the vast majority of words in written English (i.e., words that have frequencies of >10 occurrences per million) appear to bear out.

To demonstrate the power of syntactic changes on a score on the LS, Hiebert (2008) also created three alternative syntactic forms of the same text. The changes made were minor, most reflecting 21st-century grammatical conventions rather than Austen's 19th-century patterns: (a) Syntactic #1, in which two semicolons were substituted for periods, (b) Syntactic #2, in which two commas in dialogue were changed to periods, and (c) Syntactic #3, in which an unconventional sentence with no predicate but with four equivalent phrases was changed into two sentences. The shift of the LS from the original to Syntactic #3 was substantial: 300 Lexiles, or the equivalent of 1.3 grades in the high-school ranges. On the Flesch-Kincaid, the difference from the original to Syntactic #3 was 0.8 grade levels. Slight changes in punctuation meant significant reclassification on the LS, while substantial changes in vocabulary meant minor reclassification.

Another criticism that has not received substantial attention but applies to the ATOS (School Renaissance Institute, 2000) as well as to the LS is the provision of an omnibus score for a text. On the Lexile Map (MetaMetrics, 2000), *Pride and Prejudice* is given as a prototype for 1100 Lexile, and *Modern Biology* (Holt, Reinhart & Winston) for 1130 Lexile. These two texts are judged to be relatively the same in comprehensibility. An analysis of high-school biology textbooks, however, showed that they contain about 45-50% more new words than are presented in a semester of foreign language instruction (Armstrong & Collier, 1990).

Further, 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 (the unit of assessments used to obtain students' Lexile levels) that were pulled from every 1,000 words had Lexiles that ranged from 670 to 1310, with an average of 952. The range of 640 on the LS represents the span from third grade to college.



Other issues can be raised about the specific features of the LS, such as its failure to consider the frequency of the morphological family, a metric that is used in other data analyses such as the Academic Word List (AWL) (Coxhead, 2000). The failure to combine frequencies of prolific morphological families (e.g., develop, development, developing, redevelop) means that the LS of informational texts that have a high incidence of morphologically derived words will be inflated. Due to limitations of space, however, the issues that have been raised to this point are sufficient to illustrate the serious questions about interpretation and application of the LS.

## Recommendations

The LS was in the early stages of development when Chall (1985) observed, passionately and clearly, that the abuse of readability formulas in educational and publishing contexts is not inherent in the formulas themselves. While a simple response to a complex issue often results in abuse, the developer of the simple response cannot be blamed for its misuses and misinterpretations. Even so, claims have been made for the LS that were never made by Chall and Dale (1948), Flesch (1948), or Spache (1974) regarding readability formulas. Promises that the same scale can be used to match readers and texts, and claims that such matches are valid, are dubious at best. Even more than the classic readability formulas, the LS is susceptible to legitimate criticism, as was evident in the review by the NCES Panel (White & Clement, 2001). In the real world, however, teachers, publishers, and a variety of consumers need a way to sort texts. The LS is, at present, the means by which many educational units are determining relative text difficulty.

When faced with the demands to provide LS data, publishers and authors find themselves in a quandary. They have crafted specific texts for specific reasons, or have carefully developed a text with specific features for an online activity. It is highly unlikely that any readability formula will capture the distinctions that characterize these texts. I will propose several responses that are available to publishers and authors of texts for schools, including online programs.

First, the educational agencies that set mandates for text levels and review materials are likely to be more concerned when text levels are too high than when they are too low. When the typical procedures of MetaMetrics are followed, publishers submit an entire text, such as *Modern Biology* (Holt, Rinehart & Winston, 1999). In large corpora of texts such as entire textbooks or novels in which 200,000 or more words are the unit of analysis, the weight of rare vocabulary is typically less than is the case with individual sections or selections of 2,000 or 200 words. It is in the finely tuned short texts of intervention programs or of beginning reading instruction that the Lexiles will likely be variable and inflated. It is when particular sections of texts are examined—such as the introduction of a concept in a content area—that the LS will rise dramatically. As an example, when Hiebert (2008) analyzed 4,031 words from an online chemistry course (Apex Learning, 2008), the LS was 990. When an excerpt of this text (Excerpt #1 above) was analyzed, the LS was 1260. When the size of the corpus is large, it is likely that Lexile levels will have “washed” out and texts will have Lexile levels within prescribed ranges. A prediction that has yet to be tested but appears warranted from observations is that texts from online courses are likely to have lower Lexiles than their textbook counterparts because of motivational and explanatory components of these courses. These components are often written in a

journalistic style with shorter sentences and more common vocabulary—even in content areas such as chemistry.

Second, publishers can and should provide information on alternative measures of text difficulty alongside the LS information. As a result of work on Artificial Intelligence, numerous text processing procedures have been developed. In particular, the use of latent semantic analysis (LSA) means that relationships among concepts and words can be established. Landauer, Laham, and Foltz (2000) have used LSA to establish the meaning of polysemous words in specific contexts as well as synonyms within and across texts. In the Educational Text Selection software, LSA is used to predict how much readers will learn from texts based on a conceptual match between their knowledge of the topic and the content of the text (Wolfe, Schreiner, Rehder, Laham, Foltz, Kintsch, & Landauer, 1998). Another project that uses LSA, Coh-Metric (Graesser, McNamara, Louwerse, & Cai, 2004), calculates the coherence of texts on a wide range of measures. Coh-Metric is presented as a replacement for readability formulas, classic and new. While at the level of the word and not at the sophisticated level of within- and across-text coherence and cohesion, the presence of GAV (e.g., Coxhead, 2000) can be established through several text analyses programs.

There are also time-intensive systems such as the scales developed by Chall, Bissix, Conard, and Harris-Sharpley (1996) for leveling complex texts. This outstanding system represents Chall's 50 years of study on what makes a text difficult. Unlike the LS, however, in which a Lexile can be acquired for a 300,000-word text almost instantaneously (provided it is in a .txt format), the procedures of Chall et al. require individuals with expertise in students, texts, and content areas to review texts for vocabulary knowledge, sentence structure, subject-related and cultural knowledge, technical knowledge, density of ideas, and level of reasoning.

Finally, readers of this essay should not be left with the idea that I am recommending that texts be manipulated to comply with potential levels mandated by states or other educational units. It may be tempting to engage in this practice. Syntactic manipulations are so easy to make, and on the LS, they have such a strong influence. I reiterate, however, the plea that my colleagues and I made almost 25 years ago in *Becoming a Nation of Readers* (Anderson et al., 1985). Such manipulations can be pernicious to the integrity and coherence of text and create obstacles to readers' comprehension. The goal of writing and publishing texts for schools is to provide the most comprehensible text possible. When that is the goal, writers and publishers should aim to create and select the texts that are exemplars of the best knowledge identified by content area experts and researchers of language and literacy. In the end, the best evidence that a text is comprehensible is that its readers remember and apply the information in it.

## References

- Adams, M. J. (in press). The challenge of advanced texts: The interdependence of reading and learning. In E.H. Hiebert (Ed.), *Reading More, Reading Better: Are American Students Reading Enough of the Right Stuff?* New York: Guilford Press.
- 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.
- Anderson, R. C., Osborn, J., & Tierney, R. J. (Eds.). *Learning to read in American schools: Basal readers and content texts* (pp. 153-174). Hillsdale, NJ: Erlbaum.
- Apex Learning (2008). Comprehensive Courses: Chemistry. [downloaded October 20, 2008: [http://www.apexlearning.com/CourseOutlines/2008-2009/Comprehensive/chem\\_sem\\_1\\_hc.htm](http://www.apexlearning.com/CourseOutlines/2008-2009/Comprehensive/chem_sem_1_hc.htm)]
- Armstrong, J. E., & Collier, G. E. (1990). *Science in biology: An introduction*. Prospect Heights, IL: Waveland Press.
- Austen, J. (1813). *Pride and prejudice*. [downloaded on October 20, 2008 at <http://www.authorama.com/pride-and-prejudice/>]
- Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.
- Brown, W. R. (1965). Science textbook selection and the Dale-Chall formula. *School Science and Mathematics*, 65, 164-167.
- Burkey, J. E. (1954). The readability levels of recently published elementary school textbooks. Unpublished doctoral dissertation, University of Pittsburgh.
- 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. (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., & Dale, E. (1995). *Readability revisited: The new Dale-Chall readability formula*. Cambridge, MA: Brookline Books.
- Chall, J. S., Bissett, G. L., Conard, S. S., & Harris-Sharples, S. (1996). Qualitative assessment of text difficulty: A practical guide for teachers and writers. Cambridge, MA: Brookline Books.

- 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.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34 (2), 213-238.
- Cramer, W., & Dorsey, S. (1969). Science textbooks—how readable are they? *Elementary School Journal*, 70, 28-33.
- 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, and 37-54.
- Dale, E., & O'Rourke, J. (1981). *The living word vocabulary*. Chicago, Ill.: World Book/Childcraft International.
- Davison, A. & Kantor, R. N. (1982). On the failure of readability to define readable texts: A case study from adaptations. *Reading Research Quarterly*, 17, 187-209.
- Finn, P. J. (1978). Word frequency, information theory, and cloze performance: A transfer theory of processing in reading. *Reading Research Quarterly*, 13, 508-537.
- Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, 32, 221-233.
- Foorman, B. R. (in press). Text difficulty in reading assessment. In E. H. Hiebert (Ed.), *Reading More, Reading Better: Are American students reading enough of the right stuff?* New York: Guilford.
- 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.
- Hiebert, E. H. (2008). *The new readability formulas: Do they do what they claim to do?* Santa Cruz, CA: TextProject.
- Hiebert, E. H., Menon, S., & Martin, L. A. (in progress). *Digital scaffolds that support comprehension of adolescent readers*. Santa Cruz, CA: TextProject.
- Holmquist, J. B. (1968). A determination of whether the Dale-Chall readability formula may be revised to evaluate more validly the readability of high school science materials. Unpublished doctoral dissertation, Colorado State University, Fort Collins.
- Holt, Rinehart & Winston (1999). *Modern Biology*. New York: Random House Value Publishing.

- Kincaid, J. P., Fishburne, R. P., Jr., Rogers, R. L., & Chissom, B. S. (1975). Derivation of new readability formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy enlisted personnel, Research Branch Report 8-75, Millington, TN: Naval Technical Training, U.S. Naval Air Station, Memphis, TN.
- 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.
- Landauer, T. K., Laham, D., & Foltz, P. W. (2000). The Intelligent Essay Assessor. *IEEE Intelligent Systems*, 27-31.
- Lively, B. A., & Pressey, S. L. (1923). A method for measuring the vocabulary burden of textbooks. *Educational Administration and Supervision*, 9, 389-398.
- MetaMetrics. (2000). *The Lexile framework for reading*. Durham, NC: Author. [Online]. Available: [http://Lexile.com/about/\\_meta/press/21098b.htm](http://Lexile.com/about/_meta/press/21098b.htm).
- MetaMetrics (2003). Press release: National Center for Education Statistics (NCES) Utilizes Lexile Framework [downloaded October 20, 2008: <http://www.Lexile.com/News/5-21-2003.html>].
- Moore, A. C. (1935). Recoiling from reading: A consideration of the Thorndike Library. *Library Journal*, 60, 419-422.
- School Renaissance Institute (2000). The ATOS readability formula for books and how it compares to other formulas. Madison, WI: School Renaissance Institute, Inc.
- 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).
- Spache, G. (1974). A new readability formula for primary-grade reading materials. *Elementary School Journal*, 53, 410-413.
- Stenner, A. J., Burdick, H., Sanford, E. E., & Burdick, D. S. (2006). How accurate are Lexile text measures? *Journal of Applied Measurement*, 7, 307-322.
- Ta-Min, R. (2006). Comprehension of online cancer information: A propositional assessment of readability, inferences, and coherence. Unpublished dissertation, University of Waterloo (downloaded October 20, 2008 at <http://uwspace.uwaterloo.ca/etd/rtamin2006.pdf>).
- Texas Education Agency. (1990). Proclamation of the State Board of Education advertising for bids on textbooks. Austin, TX: Author.
- Walker, W. L. (1965). Readability of programmed textbooks for intermediate grades as measured by the Dale-Chall formula. Unpublished doctoral dissertation, University of Missouri.
- White, S., & Clement, J. (2001). Assessing the Lexile Framework: Results of a Panel Meeting. (Working Paper No. 2001-08). Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Wolfe, M. B. W., Schreiner, M. E., Rehder, B., Laham, D., Foltz, P. W., Kintsch, W., & Landauer, T. K. (1998).

Learning from text: Matching readers and texts by latent semantic analysis. *Discourse Processes*, 25, 309-336.

Zipf, G. K. (1935). *The psycho-biology of language: An introduction to dynamic philology*. Boston: Houghton Mifflin Company.

<sup>i</sup> A primary difference from the earlier readability formulas such as the Dale-Chall and the Lexiles, in particular, was that the semantic component was no longer accessible to users of the readability formula. While clients are told that students' performances on a Lexile-developed assessment predict reading level, it is impossible to establish what students would need to be taught in terms of vocabulary to progress to the next grade level.

<sup>ii</sup> Foorman (in press) claims that Latent Semantic Analysis (Landauer, Latham, & Foltz, 2000) has been added to the Lexile Analyzer recently. That is, the semantic space of neighboring words is used to determine the exact meaning of a word (e.g., whether *exit* is a verb or a noun, or whether *concrete* refers to building material or a non-abstract idea). Documentation of this addition to the LS is not evident on [www.Lexile.com](http://www.Lexile.com).

<sup>iii</sup> From grades 1 through 5, the spread is approximately 300 Lexile points (beginning with 100 in Grade 1). From grades 6 through 8, the spread is 250 points. In grades 9 to 12, it is approximately 230 (ranging from 200 in Grade 9 and 250 in grades 11 and 12).

<sup>iv</sup> A procedure that could be followed, in cases of highly technical text, would be to use those emanating the findings of Cohen and Steinberg (1983) and Finn (1978). That is, duplications of technical words that are core to a topic and frequently repeated within an excerpt of text could be eliminated for the Lexile analysis.