© Chris Hendrickson/Masterfile/Corbis

2

Quantitative Measures of Text Complexity

One dimension of text complexity involves quantitative measures. These primarily focus on the characteristics of the words themselves and their appearance in sentences and paragraphs. Conventional quantitative text measures do not take into account the functions of words and phrases to convey meaning, but rather focus on those elements that lend themselves to being counted, and therefore calculated. These surface structures are collectively described as readability formulas, and primarily measure semantic difficulty and sentence complexity. Gunning (2003) reports that while more than one hundred readability formulas have been developed since the 1920s, only a handful are regularly used today.

To provide a historical context for thinking about the components of readability formulas, we need to review some of the history. In 1935, Gray and Leary analyzed 228 text variables and divided them into four types: content, style, format, and organization. They could not find an easy way to

measure content, format, or organization, but they could measure variables of style. From their list of seventeen variables of style, they selected five to create a formula:

1. Average sentence length
2. Number of different hard words
3. Number of personal pronouns
4. Percentage of unique words
5. Number of prepositional phrases

Their formula had a correlation of .645 with comprehension as measured by reading tests given to eight hundred adults. These criteria have been applied to varying degrees in nearly all readability formulas since their original studies.

Word-Level Analysis

Conventional quantitative text measures do not take into account the functions of words and phrases to convey meaning, but rather focus on those elements that lend themselves to being counted, and therefore calculated.

There is a strong foundation for using quantitative measures to determine the relative level of challenge posed to a reader. The first level of analysis is at the word level. The overall length of the word suggests the degree to which a reader must decode the word, with single-syllable words considered to be easier than multisyllabic ones. As well, the frequency with which the word appears in a language supposes its familiarity to the reader. The *Brown Corpus*, developed in 1964 by Francis and Kucera at Brown University, used computational analysis of over a million words drawn from five hundred written sources, including novels, newspapers, and scientific journals, to determine each word's degree of occurrence in American English. They determined that the words *the*, *to*, and *of* collectively comprised 13 percent of the corpus, or body of words in the language. Word frequency lists used in readability formulas may number in the thousands, or even millions, but all attempt to rank-order a word's frequency of use within specific text types. The most comprehensive review of word frequency completed to date is *The Educator's Word Frequency Guide* (Zeno, Ivens, Millard, & Duvvuri, 1995), which is a listing of printed words that has been organized by how often a particular word appears in texts encountered by students at a specific grade level.

However, word frequency alone is an incomplete measure, since the context in which the word appears can increase text complexity. In order

to focus more specifically on school-aged readers, in the 1940s, Dale, later aided by O'Rourke, began developing a list of words that 80 percent of fourth graders would recognize and know. Over time, these evolved to a list of three thousand words (Chall & Dale, 1995). The genius of this work is that the researchers didn't just make a list; they applied this list as a way of determining the challenge readers might experience depending on the number of words *not* on the list. In other words (excuse the pun), a text with a higher percentage of words not among the three thousand could indicate a higher degree of complexity. Thus, a text with the words *field*, *meadow*, and *pasture* (which appear on the list) would not be deemed as difficult as a text that used the words *steppe* and *mead*, which do not appear on the list. The application of such a word list took into account what the reader might be expected to know, as well as the vocabulary demand of a word. Other word frequency lists developed since then build a corpus, or body, that is reflective of the use of a group of people, such as fourth graders or students entering high school. A key factor in this list is that Dale and O'Rourke tested and retested these words with students over a period of several decades and eventually published the list as *The Living Word Vocabulary* (1976). This sets it apart from other frequency lists.

Sentence-Level Analysis

A second level of analysis included in nearly all quantitative readability formulas is the length of the sentence. The number of words in a sentence is a proxy for several syntactic and semantic demands on a reader (e.g., prepositional phrases, dependent clauses, adjectives, and adverbs). Taken together, these press a reader's working memory to keep a multitude of concepts and connections in mind (Kintsch, 1974). Consider the following sentence from Sandra Cisneros's short story, "Eleven," about a young girl embarrassed by the shabbiness of her sweater:

> This is when I wish I wasn't eleven, because all the years
> inside of me—ten, nine, eight, seven, six, five, four,
> three, two and one—are pushing at the back of my eyes
> when I put one arm through one sleeve of the sweater
> that smells like cottage cheese, and then the other arm
> through the other and stand there with my arms apart
> like if the sweater hurts me and it does, all itchy and full
> of germs that aren't even mine. (Cisneros, 1991, p. 8)

At eighty-three words, this sentence requires the reader to process several concepts simultaneously: the sweater and its smell and feel, the clause that lists a descending sequence of numbers, the use of the word *other* to refer first to the girl's arm and then to her sleeve. An analysis of individual words alone would be insufficient; all but two appear on the Dale-Chall Word List (*itchy* and *germs* do not). We deliberately selected a long sentence to illustrate a point—sentence length can be a valid indicator of the cognitive load.

Except when it's not. Very short sentences can also tax a reader:

> For sale: Baby shoes, never worn.

Legend has it that this six-word story was written by Ernest Hemingway to settle a bar bet. All of the words appear on the Dale-Chall Word List. However, the level of inference and background knowledge needed to understand this text would challenge young readers. Readability formulas offer us a level of quantitative analysis that is not readily apparent, but should be augmented by the qualitative analyses that only a human reader can offer (Anderson, Hiebert, Scott, & Wilkinson, 1985).

We have taken time to discuss issues of word length, syllables, frequency of occurrence, and word lists because they are widely regarded as being proxies for the time needed for a reader to read the text, and the extent to which it taxes a reader's working memory (Just & Carpenter, 1992). As noted by Gunning (2003), these variables can be used as measures of semantic complexity. His insights echoed many of the dimensions described by Gray and Leary in 1935:

- Number of words not on a list of words tested and found to be known by most students at a certain grade level
- Number of words not on a list of high-frequency words
- Grade levels of the words
- Number of syllables in the words
- Number of letters in a word
- Number of different words in a selection
- Number of words having three or more syllables
- Frequency with which the words appear in print (p. 176)

Conventional Readability Formulas

Conventional readability formulas have been utilized extensively as a means to replace outdated grade-level formulas for rating text difficulty. An advantage of these readability formulas is that teachers can easily compute them using any reading material. A few of the more common formulas, and how they are used to determine readability, are reviewed next. As a way to highlight some of the differences among these, we'll analyze a passage from *The Hunger Games* (Collins, 2008). This passage (see Figure 2.1), from about the middle of the book, contains a proper noun (a character's name, *Peeta*) and some words that have been introduced previously, such as *tributes*. According to Scholastic, overall readability or the quantifiable features of the book is 5.3 grade level, but the publisher recommends the content for students in Grades 7–8.

Individual passages within the book are harder, as we will see, which means other passages must be easier. This is an important point in considering quantitative difficulty—the law of averages is at work. That does not mean that the entire text is readable just because the average suggests it is so. That said, readability formulas can be used to guide text selection in a quick and easy way. They just aren't the only guide available to teachers.

Figure 2.1 Excerpt From *The Hunger Games*

After the anthem, the tributes file back into the Training Center lobby and onto the elevators. I make sure to veer into a car that does not contain Peeta. The crowd slows our entourages of stylists and mentors and chaperones, so we have only each other for company. No one speaks. My elevator stops to deposit four tributes before I am alone and then find the doors opening on the twelfth floor. Peeta has only just stepped from his car when I slam my palms into his chest. He loses his balance and crashes into an ugly urn with fake flowers.

Source: Collins (2008, p. 134).

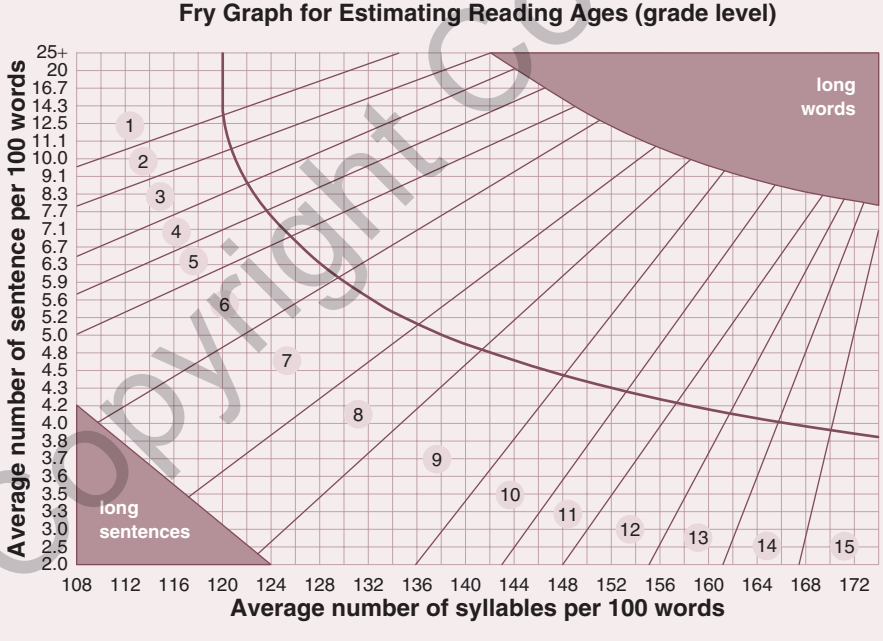
Quantitative reading formulas are notoriously unreliable on works designed for beginning readers. Hiebert and Martin (2001) note that unique characteristics of the emergent reader make issues of decodability, independent word recognition, and pattern mastery more specialized than a simple measure of readability can identify. In addition, the sentence structures for these materials may be very short, sometimes a single word, with

heavy reliance on illustrations from which the reader can draw extensive support. For these reasons, most quantitative readability formulas do not report expected measures for texts designed for very young children, primarily kindergarten and first grade. Poems, which by nature often use single words, phrases, fragments, and unconventional punctuation, also do not yield useful readability scores.

Fry Readability Formula

The primary appeal of the Fry readability formula is its ease of use, and the fact that it does not require any specialized software or hardware. Edward Fry (2002) designed this simple readability rating so that it can be calculated using the graph in Figure 2.2. The teacher selects three 100-word passages from the text, preferably one each from the beginning, middle, and end. Next, the teacher counts the number of sentences and syllables in each passage, then averages each of the two factors (number of syllables

Figure 2.2 Fry Readability Graph



Source: Fry (2002, p. 288).

and number of sentences). These two factors are then plotted on the Fry graph to yield an approximate grade level. This readability formula does not require any computer calculations, as the algorithm is embedded in the graph. For this reason, the Fry readability formula is popular among teachers who need a quick method for gaining a sense of the approximate level of difficulty. However, it does not rely on any specific vocabulary or word frequency levels, and thus can only provide limited information about a text. Using this formula, the passage from *The Hunger Games* in Figure 2.1 scored at the seventh-grade level, which is a reasonable estimate given that the content was suggested for middle school students.

Flesch-Kincaid Grade-Level Score

Another easily accessible tool for determining readability formulas can be found on the word processing software installed on your computer. Simply type in a passage from a text you would like to assess for readability, then run the calculation. For example, the Microsoft Word® program can report a Flesch-Kincaid grade-level score to approximate difficulty, using an algorithm that includes the average sentence length (ASL) and average number of syllables per word (ASW), the same elements used to calculate the Fry readability formula: $(0.39 \times \text{ASL}) + (11.8 \times \text{ASW}) - 15.59$ (Graesser, McNamara, & Louwerse, 2011, p. 42). This measure has a high correlation with the Fry graph.

The program will also report a Flesch reading-ease score by assigning the reading a number on a 100-point scale. On this scale, the higher the score, the easier it is to read. This formula is more commonly used in business to determine the difficulty of workplace documents. Both the Flesch-Kincaid and Flesch reading-ease measures calculate using the same text characteristics, but the algorithms are weighted differently to ensure that easier texts are reported as lower numbers for the grade-level purposes, and are reported as higher numbers when considering the relative ease of the text. The Flesch reading-ease score for this paragraph was a difficult 37.4, and the Flesch-Kincaid grade-level score was 12.0. The *Hunger Games* passage in Figure 2.1 received a reading-ease score of 70.7, meaning that it could be understood by students ages thirteen to fifteen, and earned a Flesch-Kincaid grade-level score of 6.9.

Advantage-TASA Open Standard

The Advantage-TASA Open Standard, a computerized readability formula more commonly called ATOS, is used by Renaissance Learning to

gauge texts used with the Accelerated Reader software. Its name reflects the partnership between Renaissance Learning (formerly Advantage Learning) and Touchstone Applied Science Associates (TASA), which developed the Degrees of Reading Power (DRP) tests. *The Educator's Word Frequency Guide* (Zeno et al., 1995) is used to determine the grade level of the words. The ATOS formula computes words per sentence, average grade level of words, and characters per word, as measured by the entire text, not just sample passages. In addition, it factors whether the text is fiction or nonfiction (the latter is considered more difficult) and the length of the book (longer texts are more difficult).

Practical advantages of the ATOS measure include the large number of trade books in its database (160,000) and the free calculation service for measuring other texts such as magazine articles and short stories. As with all readability formulas, ATOS does not take content into consideration, so *The Catcher in the Rye* (Salinger, 1951) carries a grade-level rating of 4.7. The makers caution that this measure should not be used in isolation, and each book also carries an "Interest Level" measure to further guide educators, students, and parents. Therefore, the same text has an Interest Level rating as Upper Grades (9–12). *The Hunger Games* earns a book-level measure of 5.3, but an Interest Level of Upper Middle Grades (6 and up).

Readability Formulas That Also Assess Readers

Conventional readability formulas do not factor other elements that can influence difficulty, such as content. For example, a Flesch-Kincaid grade-level analysis on a hundred-word passage from *Cat's Cradle* by Kurt Vonnegut (1998), a decidedly adult satire of a world on the brink of an apocalypse, reveals a grade-level score of 2.3 because the passage contains short, clipped dialogue. While this is not a typical result, it does highlight some shortcomings when relying on readability formulas alone without considering the content or the reader. In the 1980s and 1990s, two readability formulas were developed that attempted to account for content factors, however imperfectly, and to project estimated comprehension levels of students at each grade level. In other words, these tools can be used to assess students' reading levels and to evaluate quantitative complexity.

Degrees of Reading Power

This widely used formula uses "sentence length, number of words not on an updated version of the Dale list, and average number of letters per word"

(Gunning, 2003, p. 178). A review by Graesser, McNamara, and Louwerse (2011) found that both DRP and Lexile (discussed in the following section) correlated strongly with the Flesch-Kincaid readability measure. Gunning (2003) reports that DRP (Koslin, Zeno, & Koslin, 1987) uses a variation of an older readability formula called the Bormuth. What sets the Bormuth apart is that it was the first formula to use cloze as a criterion measure. While the DRP formula is proprietary, the Bormuth formula uses average word length (AWL), average sentence length (ASL), and average number of familiar words (AFW, defined as those words that appear on the Dale-Chall list of three thousand words) as follows:

$$\text{Bormuth Readability Score} = 0.886593 - (\text{AWL} \times 0.03640) + (\text{AFW} \times 0.161911) - (\text{ASL} \times 0.21401) - (\text{ASL} \times 0.000577) - (\text{ASL} \times 0.000005)$$

An advantage of DRP is that it calculates a reader's performance with text using the same scale so that educators can match readers and books. DRP does not make readability scores of assessed texts publicly available, so we are unable to report the DRP level for the *Hunger Games* passage in Figure 2.1.

TextEvaluator

Originally developed as SourceRater as a tool to select passages for use on assessments, TextEvaluator provides a single, overall measure of text complexity using a scale that ranges from 100 (appropriate for extremely young readers) to 2,000 (appropriate for college graduates). This is similar to the scales used in other tools, including Lexile. A unique feature of TextEvaluator is that it also produces information about text variation and which of the eight factors may contribute to the complexity. Some of these factors are familiar (e.g., academic vocabulary, word unfamiliarity, syntactic complexity), but some are less so, including the following:

- **Concreteness** measures the number of words that evoke clear and meaningful mental images as they are likely to be less difficult than those that do not.
- **Lexical cohesion** measures the likelihood that the text will be seen as a "coherent message" compared with a collection of unrelated clauses and sentences.
- **Level of argumentation** measures the ease or difficulty of inferring connections across sentences when the text is argumentative.

- **Degree of narrativity** measures the features that indicate it is more characteristic of narrative than nonnarrative or expository writing.
- **Interactive/conversational style** measures the degree of conversational style.

Reading Maturity Metric

At this time, the Reading Maturity Metric is in beta testing by Pearson publishers. It's an appealing tool because it relies on word maturity, or the ways in which meanings of words and passages change as learners develop literacy skills (Landauer, Kireyev, & Panaccione, 2011). As an example, consider the word *trust*. A younger person may know the word as it relates to confidence. A person with more word maturity also knows that it can be a type of organization, often with funds associated with it. Thus the phrase *trust baby* is unclear without the context, and the Reading Maturity Metric is being tested to take into account the sophistication of a reader's word knowledge.

Lexile

Each of the tools we have discussed has aligned with grade-level equivalents, and each provides a range for reading proficiency, not a specific and exact target that must be met.

This commercially available readability formula, developed by Smith, Stenner, Horabin, and Smith (1989), is used widely by textbook and trade publishers and testing companies to designate relative text difficulty among products. For example, the National Assessment of Educational Progress (NAEP) and the Programme for International Student Assessment (PISA) both use Lexile. Like DRP, the Lexile scale relies on a 2,000-point scale that is used to describe both readers and text, making it easier for teachers to match one to the other. The Lexile scale score assigned to *The Hunger Games* is 810, which means that it would be of appropriate reading difficulty for students in fourth or fifth grade. As we have noted, however, many of the themes in the book are not appropriate for students at this grade level.

Both DRP and the Lexile scale rely on conventional text analysis algorithms, with one notable exception: they can be used to assess students in order to pair texts with readers. Both measures apply a similar approach to assessing students, using cloze items within reading passages. By using the same scale, a teacher can match a student's DRP or Lexile scale score with a text at that same level. Additionally, teachers can use information about a reader's quantitative score to identify texts that appropriately challenge him or her.

The readability formulas discussed in this chapter thus far vary somewhat in their algorithms and the factors they use to quantify a text. These formulas draw on characteristics that serve as approximations of overall difficulty: length of word, frequency of occurrence in the language, number of syllables, sentence length, or inclusion of words on a specific word list, such as the Dale-Chall list. Some of these formulas are better than others at predicting comprehension. We present a number of different formulas because each is used, to a varying degree, in school systems, and thus informed practitioners should understand what the formula is measuring and what it is not measuring. That said, most of the formulas account for about 50 percent of the variation in comprehension. The Lexile formula is better, predicting about 75 percent of the variation (see Smith, Stenner, Horabin, & Smith, 1989).

Each of the tools we have discussed thus far has aligned with grade-level equivalents (see Figure 2.3). Note that there is overlap across the grades, meaning that the upper end of one grade will likely not begin the next grade. Each tool provides a range for reading proficiency, not a specific and exact target that must be met.

Figure 2.3 Common Scale for Band Level Text Difficulty Ranges

Common Scale for Band	Text Analyzer Tools					
	ATOS	DRP	FK	Lexile	RM	SR
2nd–3rd	2.75–5.14	42–54	1.98–5.34	420–820	3.53–6.13	0.05–2.48
4th–5th	4.97–7.03	52–60	4.51–7.73	740–1010	5.42–7.92	0.84–5.75
6th–8th	7.00–9.98	57–67	6.51–10.34	925–1185	7.04–9.57	4.11–10.66
9th–10th	9.67–12.01	62–72	8.32–12.12	1050–1335	8.41–10.81	9.02–13.93
11th–CCR	11.20–14.10	67–74	10.34–14.20	1185–1385	9.57–12.00	12.30–14.50

Source: National Governors Association & Council of Chief State School Officers (n.d.).

Key:

ATOS = ATOS® (Renaissance Learning)
DRP = Degrees of Reading Power® (Questar Assessment, Inc.)
FK = Flesch Kincaid® (public domain, no mass analyzer tool available)
Lexile = Lexile Framework® (MetaMetrics)
RM = Pearson Reading Maturity Metric© (Pearson Education)
SR = SourceRater© (Educational Testing Service)

Measures not in concordance table:

REAP (Carnegie Mellon University)
Coh-Metrix (University of Memphis)

Measuring Coreference and Cohesion

One concern raised by some educators and reading researchers is that coherence and cohesion primarily measure surface-level complexity but do not get at deeper levels of meaning that are necessary to read longer and more sophisticated texts (Davison & Kantor, 1982). A newer readability measure, called Coh-Metrix, attempts to look at the deeper structures of a text, especially in its ability to present ideas coherently. These analyses have primarily been within the fields of linguistics, artificial intelligence, and computational linguistics (Graesser, McNamara, & Louwerse, 2011). With the advancement of newer tools of analysis, especially those that can parse texts at a fine-grained level and those that account for cohesion (the relationship between given and new knowledge), a more complex method of computing readability includes aspects of both semantic and syntactic features.

One such tool, latent semantic analysis (LSA), which Coh-Metrix uses, offers a way to move beyond surface-level measures of word and sentence length and word frequency, to mathematically measure how words and phrases connect with other words and phrases across a text (Landauer, McNamara, Dennis, & Kinstch, 2007). This measure also takes into account the amount of implicit knowledge needed to understand the relationships between words and ideas. For instance, the word *cup* is associated with other words such as *fork* and *plate*, as well as *coffee*, *set the table*, and *wash the dishes*, even though these terms may not appear in the text. An LSA analysis forms a map, or matrix, of connections that are beyond a human's ability to detect and measure.

In addition to LSA measures, computational linguistics researchers have sought to further quantify other elements of text, including parts of speech, genre of the text, psycholinguistic dimensions of words such as relative level of abstraction or concreteness, and propositional density (how a noun phrase is linked to an agent, recipient, location, or object). These and other related measures work together to influence a text's *coreference* (the extent to which a word or phrase references similar words or phrases in the surrounding text). For instance, argument overlap "is the most robust measure of lexical coreferentiality in that it measures how often two sentences share common arguments (nouns, pronouns, and noun phrases)" (Crossley, Dufty, McCarthy, & McNamara, 2007, p. 199). As a simple example, consider these two sentences:

> The bookshelves sagged under the weight of the heavy, dusty books. No one had checked out these books from the library for many years, as evidenced by the infrequent checkout dates on the cards in their lonely front pockets.

The word *library* locates the bookshelves and the books, and the word *their* coreferences *books* in both the first and second sentences. *Checked out* and *checkout* are different parts of speech, but they coreference each other due to proximity as well as agency. A latent semantic analysis of the same short passage might reveal that there are further relationships beyond the text, including *librarian*, *library card*, and the process needed to borrow and return a book. Taken together, latent semantic analysis, psycholinguistic measures, and coreferencing combine to contribute to a text's cohesion—that is, the number of its meaning relations. We are speaking not of the overarching meanings related to theme, main ideas, and so on, but rather of the way syntax and semantics interact to develop a coherent message within and across sentences and paragraphs within the same text. This tool uses sixty-four indices that report on measures, or metrics, related to the research on discourse, language, and cognition to assign text difficulty.

The Coh-Metrix tool is available at no cost at http://cohmetrix.com. Perhaps the most valuable application of Coh-Metrix is in the authors' recommendations about the identification of texts across five dimensions, each addressing a specific purpose and reader:

1. **Challenging Texts With Associated Explanations.** Some assigned texts are considerably beyond students' ability level. In such cases, students need comments by a teacher, tutor, group, or computer that explain technical vocabulary and points of difficulty. Students are greatly stretched by exposure to difficult content, strategies, and associated explanations.
2. **Texts at the Zone of Proximal Development.** Some assigned texts are slightly above the difficulty level that students can handle. These texts gently push the envelope—they are not too easy or too difficult, but just right.
3. **Easy Texts to Build Self-Efficacy.** Easy texts are assigned to build reading fluency and self-efficacy. Struggling readers can lose

self-confidence, self-efficacy, and motivation when beset with a high density of texts that they can barely handle, if at all.

4. **A Balanced Diet of Texts at Varying Difficulty.** Texts may be assigned according to a distribution of alternatives 1, 2, and 3 above, mostly in the zone of proximal development. The balanced diet benefits from exposure to challenging texts, texts that gradually push the envelope, and texts that build self-efficacy. This approach also includes texts in different genres.

5. **Texts Tailored to Develop Particular Reading Components.** Texts may be assigned adaptively in a manner that is sensitive to the student's complex profile of reading components. The texts attempt to rectify particular reading deficits or to advance particular reading skills. (Graesser, McNamara, & Kulikowich, 2011, p. 232)

These recommendations challenge us to apply quantitative measures in ways that create a text gradient that not only considers the reading itself but also takes the reader and the learning context into account. This, however, has not been seen as the primary function of quantitative reading formulas, and their use and misuse has resulted in cautions and criticisms.

Cautions About Quantitative Analysis of Text

As we have noted, quantitative measures used in isolation can result in inappropriate content being assigned to students. That doesn't mean that these tools and their resulting data are useless; rather, they have to be interpreted and used in conjunction with qualitative factors of text complexity. It is misguided and problematic to demand that teachers use texts that fall only within the quantitative text range associated with a specific tool. Text selection is multidimensional, not unidimensional. Further, selecting texts based only on the quantitative score is a form of censorship, blocking students from information and ideas that they may want to explore, simply because of the words and sentence structures the author used.

Another criticism of quantitative reading formulas is that they have been used as a device to manipulate text to meet a fixed numerical value, regardless of its effect on the text itself. For example, publishers may remove or

substitute words or phrases in order to lower the quantitative score, but in the process inadvertently make the reading more difficult to understand. For example, signal words such as *first* and *last*, and transitional phrases like *in conclusion*, add length to the sentence and can thereby raise the score. But words like these actually assist the reader by helping him or her internally organize the information. In fact, from a qualitative perspective (discussed further in Chapters 3 and 4), a profusion of signal words alerts the teacher to the fact that a text is using an internal structure to scaffold the reader's understanding. Conversely, their removal can lower the readability score but end up making the text far less coherent. Higher readability scores do not automatically signal difficulty. Beck, McKeown, Omanson, and Pople (1984) demonstrated that texts with higher levels of coherence and vocabulary were easier to comprehend than similar texts that had been stripped of these features.

Selecting texts based only on the quantitative score is a form of censorship, blocking students from information and ideas that they may want to explore, simply because of the words and sentence structures the author used.

Informational texts that use a high degree of technical vocabulary may score much higher due to the relative rarity in the general corpus, with no way to account for their much more frequent use in a specific text. For example, science texts have technical vocabulary that is comparatively rare on word frequency lists but is commonly used within a discipline (Cohen & Steinberg, 1983). The word *photosynthesis* is rare when compared to all words, but much more frequent in the field of life sciences. Therefore, a biology textbook might have a higher readability score due to the presence of such a word, despite the fact that the topic is deeply explored within its pages. *The Educator's Word Frequency Guide* (Zeno et al., 1995) was developed in part to address these discipline-specific concerns, and provides an "index of dispersion" for the use of such words.

Conclusion

Since the early part of the 20th century, educators have sought ways to order or level text through quantitative measures of readability. These formulas vary somewhat, but primarily measure surface-level features of a text, especially focusing on word and sentence length and frequency of word occurrence on a generated list. More recent developments utilize the availability of digitized texts to analyze longer texts, not just samples. Most important, advances in computational linguistics, psychology, and artificial intelligence have opened the door to a new generation of analytic tools that

provide a more fine-grained measure of the relationships of words to one another and the mental models that are necessary to understand them. A summary of the readability formulas discussed in this chapter can be found in Figure 2.4.

As with all measures, each can report accurately on some aspects, while other equally important elements remain untouched. For this reason, quantitative measures should be viewed as an important step, but by no means a final one or necessarily the first one, in determining the optimum text for a reader. Readability, after all, should never be confused with reading ability. In the next chapters, we explore another necessary element for determining text complexity: qualitative analysis of literary and informational texts.

Figure 2.4 Summary of Quantitative Text Measures

Name	Purpose	Factors Used	Ease of Use	Notes
Fry Readability Formula	Assesses text difficulty	Sentence length and syllables	Easy; use graph	Primary–college
Flesch-Kincaid Grade-Level Score	Assesses text difficulty	Sentence length and syllables	Easy; use word processing software	K–12
Flesch Reading-Ease Score	Assesses text difficulty	Sentence length and syllables	Easy; use word processing software	Reports relative ease as compared to students in Grade 5–college
Advantage-TASA Open Standard (ATOS)	Assesses text difficulty	Words per sentence, grade level of words, and character length across entire text	Easy; free online calculator and extensive published booklist	Factors fiction/nonfiction and length of text
Degrees of Reading Power (DRP)	Assesses text difficulty and reader skills using same scale	Sentence length and relative word frequency	Hard; proprietary software	Designed as criterion-referenced measures for use in Grades 1–12
TextEvaluator	Assesses text difficulty and identifies problematic areas	Vocabulary and sentence structures	Hard; uploaded text must meet all requirements	Considers a range of factors that impact comprehension
Lexile Scale	Assesses text difficulty and reader skills using same scale	Sentence length and relative word frequency	Hard; proprietary software Easy; searchable database	Reports as grade bands (Grades 2+), uses a similar scale to report student reading ability as measured by cloze items
Coh-Metrix	Assesses texts on 64 indices, including measures of text cohesion, linguistic elements, and parsers	Parsers, propositions, and latent semantic analysis, as well as traditional readability measures	Easy; use online calculator	Reports require a high degree of technical knowledge to interpret