Text Complexity

By Dr. Timothy Shanahan

Distinguished Professor Emeritus, University of Illinois at Chicago

Overview

All texts are not equal. Some are harder to read and comprehend. Just as there are differences in children, there are differences in the texts that we ask children to read. Some children read better than others, and there are a variety of reasons for these differences in reading abilities. Similarly, some texts are easier or harder, and there are several reasons for these differences.

What Makes Text Complex?

Since the 1920s there has been interest in measuring or predicting text difficulty or readability (Lively & Pressey, 1923). The idea of readability has been to try to array texts along a continuum, from easiest to hardest. The most certain way to do this would be to have people read the texts and complete comprehension tests for each one. The texts that most people understood would be rated as easy, and those that fewer people grasped would be the harder. But readability estimates are not quite so direct.

Of course, it would be too costly and inconvenient to test every text in that way, so the idea emerged that sound readability measurements should be based on the features of the texts themselves. That way someone could count up the types of words, sentences, and other text features, and then use these counts to predict which texts would give readers the most difficulty. Readability research has primarily focused on trying to come up with ways of enumerating and summarizing text features so that they lead to accurate predictions; in other words, a readability measure would sequence texts in the same order that a series of comprehension tests would.

Readability measures have evolved since they first appeared. Now, most readability formulas only include two factors: word sophistication and sentence complexity (Klare, 1984). Thus, measuring the readability or complexity of a text involves, first, an evaluation of the words: counting the average numbers of letters or syllables, checking on the frequency of the words (common words tend to be easier than rare words), or considering how abstract or concrete the words may be.

Then, an estimate of the sentence complexity is needed; perhaps average sentence lengths (longer sentences often are more complicated than shorter ones).

Finally, these word and sentence measures combined into mathematical formula used to predict actual reading comprehension. Once the measure has proven to sequence texts in the right order—presumably in the same sequence that would result from a series of reading comprehension tests.

Originally, readability measures expressed text levels in terms of school grade levels. For instance, the readability of a text might be stated as being at a third-grade or fifth-grade level. These days most measures use other those mysterious numbers into grade levels. What that means is that these graded readability scores can be interpreted as a prediction that the average student in that grade level would be able to read the text with understanding; saying that a book is at the fourth-grade level means, not that all fourth-graders could read such a book, but that the average ones should be able to.

The first such readability formulas improved our ability to predict text difficulty, but not by very much (they could account for only about 35% of the variation in actual reading comprehension performance). Over time, they improved, and these days most of them can account for about 50-60% of the comprehension variance, and the various best are now able to explain about 90% of reading comprehension differences (Smith, 2012). That means that these ratings aren't perfect—sometimes a text will be easier or harder than predicted—but they are pretty darn good.

If they are so good, why are they wrong sometimes? There are two basic reasons for this. The first has to do with a basic limitation of the measures of themselves. Remember, only word and sentence difficulty is being considered. We know there are other text features that can impact comprehension, but these other complexity features are not accounted for directly in readability. These measures do well because those unmeasured, text features tend to be correlated with the word and sentence challenges that are included. Texts that have simple wordsand sentences will usually be about simpler concepts and will use simpler or more straightforward organizations. It is possible, however, for an author to play a bit of a trick in this regard. In spite of their seemingly simple words and sentences, some authors might sneak in some complication that plays havoc with the prediction. It is possible to count up additional text features, but that approach doesn't necessarily improve accuracy much and it reduces efficiency.

What are some of these other text complexity factors – beyond sentence complexity and vocabulary sophistication – that are likely to influence reading comprehension? One factor is the complexity of the ideas themselves. Some topics (e.g., rocket science, brain surgery, the infield fly rule), are more complicated, sophisticated, or subtle than others. Also, ideas connect with each other across a text through a text property called coherence. Readers have to interpret pronouns, synonyms, and other connectors to make sense of a text. The greater the distance across ideas, the more varied the connections, and the more competition for a link (if the pronoun "he" is mentioned, how many male characters could be the topic of discussion?), the harder the text will be. Text structure or organization matters, too. Ideas can be arranged across text in many ways (e.g., time sequence, causation, lists, comparison), sometimes in ways that can be pretty complicated. Some texts may be more-or-less complete, including sufficient background information and context to help readers to interpret the ideas; other, more difficult to understand texts, may require that readers themselves bring this kind of information to the text. Finally, texts may use various literary devices (e.g., irony, repetition, metaphor) or data presentation tools (e.g., tables and charts), adding to the interpretive burden.

Another reason for the imperfect measurement of readability formulas has to do with what we are trying to predict: reading comprehension. Readers differ in what they bring to a text in background knowledge, motivation, abilities to cope with ambiguity, and so on. Even if two students have the same "reading level" they might perform differently with a particular text, which would obviously reduce the accuracy of the text complexity prediction. Reading comprehension will always be the product of two factors: the text and the reader. Readability only measures the variations in the text; reader variations can present unpredictable factors that a readability measure can never capture specifically.

Relationship of Text Complexity and Learning

Readability formulas predict reading comprehension on the basis of text features. But what connection does text complexity have with learning to read? Scholars have long claimed that the match of text difficulty and a reader's ability to read will be an important determinant of learning (Betts, 1946).

The original idea was that it is important to place students in texts neither too easy nor too hard. If the students can read a text easily without help from a teacher or someone else, then there would be little to learn from that text. If the students struggle too much to make sense of a text—that is the text is relatively too hard for them, there would be a lot to learn, but the students could be too overwhelmed or frustrated to learn—even with the help of a good teacher. From this basic conception, came the idea that there are independent, instructional, and frustration levels of text placement. Independent level refers to texts that students can read successfully without help; frustration level refers to those texts that too hard for students to learn from; and instructional level would be the optimum level that provided both an opportunity for learning, but without too much possibility of frustration.

As reasonable as this basic notion of an instructional level may seem, actually determining which texts students would learn most from requires some knotty measurements. The most direct way to study this problem would be to identify children at the same reading level and then to randomly assign them to reading instruction using texts written at different levels. Some groups would be taught from books that were easy for them, while others would be taught from relatively harder texts. The results of such experiments would reveal which student/text match leads to the greatest amounts of learning.

Unfortunately, there have been few such studies and too many teachers have been willing to rely upon authoritative assertions on this matter, rather than empirical evidence. The most influential claims have indicated that students need to be taught from texts they can read with 95-98% oral reading accuracy and with a reading comprehension of 75-89% (Betts, 1946).

Recent claims for optimum text complexity levels have been based on minor adjustments to these criteria, despite the fact that we now know Betts' made those numbers up without evaluating them against student learning (Shanahan, 1983).

References

Betts, E. A. (1946). *Foundations of reading instruction.* New York: American Book Company.

Klare, G. R. (1984). *Readability. In P. D. Pearson, Handbook of reading research* (pp. 681-744). New York:

Longman. Lively, B. A., & Pressey, S. L. (1923). *A method for measuring the 'vocabulary burden' of textbooks.*

Educational

Administration and Supervision, 9. 389-398.t

Morgan, A., Wilcox, B. R., & Eldredge, J. L. (2000). *Effect of difficulty levels on second-grade delayed readers using dyad*

reading. Journal of Educational Research, 94, 113-119.

Shanahan, T. (1983). *The informal reading inventory and the instructional level: The study that never took*

place. In L. Gentile, M.

L. Kamil, & J. Blanchard (Eds.), Reading research revisited (pp. 577-580). Columbus OH: Merrill.

Shanahan, T., Fisher, D., & Fey, N. (2012). *The challenge of challenging text.* Educational Leadership, 69(6),

58-62. Smith, M. (2012). *Not so common: Comparing Lexile measures with the standards' other text*

complexity tools. Raleigh, NC:

MetaMetrics.

Venezky, R. L., & Johnson, D. (1973). *Development of two-letter sound patterns in grades one through three.* Journal of Educational Psychology, 64, 109-115.