

The mechanism used by readability formulas makes them unreliable.

Readability Formulas in the New Millennium: What's the Use?

Karen A. Schriver
KSA Document Design & Research, Inc.
33 Potomac Street
Oakmont, PA 15139
ks0e@andrew.cmu.edu

Abstract

While readability formulas were intended as a quick benchmark for indexing readability, they are inherently unreliable: they depend on criterion (calibration) passages too short to reflect cohesiveness, too varied to support between-formula comparisons, and too text-oriented to account for the effects of lists, enumerated sequences, and tables on text comprehension. But readability formulas did spark decades of research on what comprehension really involves.

1.7.5 Document analysis—human factors

Keywords: *validity, criterion passages, reliability*

George Klare, one of the godfathers of readability, argued repeatedly over the years that readability formulas are of practical use. Reprinted for this classic series, Klare's 1963 text is one of his earliest in support of readability. In this chapter, Klare carefully positions what readability formulas are good for and what they are not. Like his other writings, this one invites the reader to agree on a basic premise: Despite the limitations of readability formulas, they should have a place in our thinking as communicators. In his seminal article on readability in 1984, Klare reiterates this perspective by telling us that, if the formulas are properly applied and their limitations kept in mind, they can "provide quantitative, objective estimates of reading difficulty" (p. 684). Klare and others who explored the uses of readability formulas between the 1950s and 1980s published dozens

of articles to point out their potential utility, most all of them ending with a call for more research that would lead to better formulas.

The call for new and better formulas, however, went largely unheeded as readability fell into disfavor among researchers and theorists who studied reading and writing processes. Although there were criticisms of the formulas as early as the 1930s, it would be in the late 1970s through the mid-1980s that attacks on readability formulas proliferated (for example, Davison, 1984). With titles such as "Readability Is a Four-Letter Word" (Selzer, 1981) and "Readability Formulas: What's the Use?" (Duffy, 1985), authors aggressively challenged the utility of the formulas. (I reuse Tom Duffy's subtitle for this commentary.)

Criticisms Of The Reliability And Validity Of Readability Formulas

Duffy and his colleagues (1982, 1985), for example, pointed out serious flaws in the reliability and validity of the formulas. Their primary criticism was that the predictor variables on which the formulas are built (i.e., sentence length and word frequency) are not the best predictors of comprehension. In short, the formulas lacked face validity (they do not measure what they claim to measure, that is, human comprehension). The formulas fail to consider the range of components that a valid and reliable model of comprehension must account for—reading skill, subject-

matter knowledge, motivation for reading, genre being read, context for the reading, and purpose for reading (see Schriver, 1989).

Advocates of the formulas countered such criticisms by arguing they never claimed readability formulas accounted for all these components. Advocates stressed that the formulas were intended as a quick objective benchmark for indexing readability. But herein lies the problem. Although the formulas seem objective because they provide a numerical score, they are inherently subjective because the score they generate is derived by comparing the text under analysis to a standard text (called a criterion passage) that may bear little resemblance to the text under analysis.

As Klare (1984) himself points out, most criterion passages were "around 100 words and seldom as long as 250 words in length" (p. 708). Not only were the passages too short to measure complex issues such as cohesiveness, but the criterion passages varied from formula to formula, making it difficult to compare the results of one formula with another. In other words, because the baseline for defining "very easy" to "very difficult" shifts from formula to formula, running several formulas on the same text may lead to different scores. Some researchers suggest that the scores generated by different formulas on the same text can vary as much as three grade levels (see Harrison, 1980). Consequently, if organizations employ the formulas for sorting out which documents to revise, they may be making decisions not on objective advice about how good the texts are, but on the noise generated in part by the variability of the criterion passages.

And alas, characteristics of the criterion passages are not the only source of noise in the scores. Worse than not necessarily being able to obtain the same score from one formula to the next is that users can learn to manipulate their score by changing text features that do not contribute to comprehension. To see why, we

... the formulas were intended as a quick objective benchmark for indexing readability. But...they are inherently subjective.

need look no further than to how the formulas operate.

Readability formulas act as algorithms that count the number of syllables per word and the number of words per sentence. One feature of these algorithms is that they rely on periods at the ends of sentences to determine when to stop counting. This effectively defines a sentence as the number of words between one period and the next. Consequently, the formulas penalize text elements that do not use periods. If text elements such as lists or numbered sequences do not employ periods at the end of each item, the count keeps running, and the score "gets worse" as the number of words increases before reaching the next period.

The algorithm would work if writers would "simply behave" and stop using itemized lists and enumerated sequences that do not require final punctuation. But writers who create technical, scientific, and medical documents employ these devices in an effort to meet rhetorical goals such as concision, focus, and parallelism. The content presented in these lists and numbered sequences is not peripheral to the text's meaning.

Oddly, some advocates of the formulas argue that writers should simply remove all itemized lists and numbered sequences from the text before they run the formulas. But writers might as well not use the formulas if they have to "doctor the text" before diagnosing what's wrong with it. We can imagine that ornery writers could improve their text's score simply by adding periods. For added dramatic flourish, they could scramble the words of each sentence, turn the typography white and the background behind the text light yellow, run the formula again, and generate the same inflated score.

The Legacy: Writing To The Formula

What happened in organizations that encouraged the use of readability formulas is that writers found ways to "write to the formulas"

in order to get a better score. Since all one needs to do is add more periods, writing skill is not even required. I have seen this strategy in some Japanese companies that do not have a history of working with professional technical communicators. With limits placed on the number of words per sentence they can use, writers have a clear aim: produce short sentences. Of course, Klare was one the first to point out that the formulas should not be used for producing text. A nice goal, but with the "seducible moment" built-in from the start, "writing to the formula" was inevitable.

A different but well-discussed limitation of the formulas is that they do not account for graphic features of text. Built on a text-centric view of understandability, the formulas happily ignore tables, charts, graphs, and pictures. Moreover, they cannot account for the dynamic ways in which prose and graphics interact: complementary, supplementary, redundant, stage-setting, and juxtapositional (for a discussion of these ways, see Schriver, 1997). These omissions spell disaster for analyzing today's popular visually oriented genre: the Web page.

Postscript

These limitations regarding the reliability and validity of readability formulas leave us with the nagging question: Has there been an upside to the use of readability formulas? Two things suggest yes.

First, the formulas serve to remind people who would otherwise be unaware of issues of readability to consider them. In some very important areas such as health literacy, a large percentage of the documents produced are created by people without training in professional writing. For these novice professional writers, the formulas offer a reasonable threshold to aim for. They are helpful, even if in a limited way.

Second, the formulas have served the very useful function of igniting debate among reading and writing researchers. The formulas catalyzed an enormous body of research about what

comprehension means. We rarely acknowledge it, but the formulas also helped catapult the development of usability testing methods. Uncovering the limitations of readability formulas had the concomitant benefit of helping us make discoveries about the processes of reading and writing. Recognizing the drawbacks of that early work led us to set new goals for theory and research, broadening our attention from readability to understandability and usability. For opening our eyes to these possibilities, we can thank scholars such as George Klare. They directed our attention to key issues and defogged the lenses with which we look.

References

- Davison, A. (1984). Readability—Appraising text difficulty. In R. C. Anderson, J. Osborn, and R. J. Tierney (eds.), *Learning to read in American schools: Basal readers and content texts* (pp. 121-151). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Duffy, T. M. and Kabance, P. (1982). Testing a readable approach to text revision. *Journal of Educational Psychology*, 74, 733-748.
- Duffy, T. M. (1985). Readability formulas: What's the use? In T. M. Duffy and R. Waller (eds.), *Designing usable texts* (pp. 113-143). Orlando, FL: Academic Press.
- Harrison, C. (1980). *Readability in the classroom*. Cambridge: Cambridge Univ. Press.
- Klare, George R. (1984). Readability. In P. D. Pearson, R. Barr, M. L. Kamil, and P. Mosenthal (eds.), *Handbook of reading research* (pp. 681-744). NY: Longman.
- Rothkopf, Ernest Z. (1985). Writing to teach and reading to learn: A perspective on the psychology of written instruction. In H. Singer and R. B. Ruddell (eds.), *Theoretical models and processes of reading*, 3rd ed. (pp. 879-907). Newark, DE: International Reading Association.
- Schriver, K. A. (1989). Evaluating the quality of text: The continuum from text-focused to reader-focused methods. *IEEE Transactions in Professional Communication*, 32(4), 238-255.
- Schriver, K. A. (1997). *Dynamics in document design: Creating texts for readers*. NY: Wiley.
- Selzer, J. (1981). Readability is a four-letter word. *Journal of Business Communication*, 18, 23-34.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage, and that all copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers, or to redistribute to lists, requires prior specific permission and/or a fee. © 2000 ACM 1527-6805/00/08—0138 \$5.00