Striking Out with Readability Formulas

By Tom Lang, MA
Tom Lang Communications and Training

The Opening Pitch

Readability formulas have been batted around for more than 50 years. Jessica Ancker's article in this issue, "Developing the Informed Consent Form: A Review of the Readability Literature and an Experiment" (*page 97*), nicely summarizes their strengths and weaknesses.[1] Many people believe that readability formulas have merit.[2,3] However, thoughts that these formulas can accurately measure "readability" or that "readability" can be indicated with a single number are way off base. The thesis of this editorial, then, is that readability formulas are a wild pitch: they do more harm than good when used to revise a text. Their one, potentially positive use is to help writers learn more about the structure of English.

Strike 1

There are dozens of readability formulas, each just about long enough for a PhD dissertation.[4,5] However, the notion that the so-called readability of any written document could be accurately characterized by a single number was always, and remains, delusional, sort of like the Cubs winning the pennant. This single number is the "reading grade level," a number between 1 and usually 20, referring to grades 1 through 12 and then up to 8 years of post-secondary education. (The reading grade level is 1 of 2 common measures of readability produced by readability formulas, the other being a scale from 0 to 100 indicating the relative "difficulty" of reading different texts.)

What does a "reading grade level" mean, really? It is generally interpreted to mean that the text can be understood by individuals with at least the same number of years of formal public education. Thus, a document written at, say, the 5th-grade level is supposed to be understandable by 5th graders or by those who have completed a 5th-grade education. But the reality is different when you consider how reading grade level is determined.

Researchers begin with a large number of short texts written at various "levels of complexity" in terms of grammatical structure, number of main and subordinate ideas, vocabulary, and so on. These texts are given to a variety of readers (often school children), who are then tested for their understanding of the text. Those texts for which half the readers of a given scholastic grade answer half the questions correctly are identified as the normative texts for that grade. *Strike 1: Even if reading grade level were a valid concept, by definition, technically only half of readers would understand half of what is in the text written at a given grade level.*

Strike 2

Once these normative texts have been identified for each grade level, the measurable characteristics of each text are determined. Usually, these characteristics are easily countable things, such as average sentence length in words, average number of syllables per word, number of words that appear (or not) on standardized vocabulary lists, and so on. More sophisticated formulas use more sophisticated measures, such as ratio of concrete to abstract nouns, number of prepositional phrases, percentage of sentences in the passive voice, and so on. (The most sophisticated research—which is not concerned with readability formulas *per se*—uses the most sophisticated measures, such as "propositional density," the number and type of "transformations" of a given core sentence, or the number and nature of "schemas" used for organizing perception.[5])

In any event, the characteristics of the normative texts are fed into regression equations that eventually result in a mathematical model (the readability formula) that will predict the "reading grade level" from a weighted combination of the characteristics included in the model. The assumption of this process is that every text with the same reading grade level is assumed to be understandable by the same group of readers; those with the required level of education—irrespective of content complexity, reader experience or motivation, graphic design features, figures, tables, and anything else that isn't easily countable, including word order. *Strike 2: It makes no sense to assume that all texts with the same combination of characteristics as the normative text or with the same readability score are equally "readable."*

The issue of word order alone allows me to eject readability scores from the writing game with a simple illustration. The two sentences on the next page have the same readability scores, but one is coherent and the other is not because the words are in random order:

"When we clarify meaning, sometimes we coincidentally reveal it."

"It reveal we when sometimes coincidentally meaning, clarify we."

Strike 3. You're Out!

Research conducted on a sample of a population can be generalized to the entire population only if the sample is representative of that population. In the same way, readability formulas normed on one population may have no validity when applied to texts for readers of an entirely different population. So, on whom were these formulas normed? Historically, the readers were usually middle class white people who lived near universities and their children, who went to public schools. Some, like the Flesch-Kincaid Grade Level scale, were normed on US Navy personnel, the overwhelming majority of whom were young men.

A related concern is when the formulas were last normed. Many were normed in the 1940s and 1950s; the most recent date I found was the early 1970s. Has literacy improved, declined, or remained the same since the 1940s? Remember that each formula has its own normative population and is updated on its own schedule, if it is, in fact, updated at all. *Strike 3: Unless you know who was in the normative population and when the formulas were normed, you don't know to whom the formulas actually apply.*

The Practice Pitch

As near as I can tell, the only positive use of readability formulas is that some computerized versions available in word processing programs flag common writing problems. For example, the program may underline all instances of the passive voice, which allows writers to learn what the passive voice is and to change it to the active voice if they so choose. Such a program may also flag nominalizations (verbs that have been turned into nouns), double negatives, unusually long sentences, incomplete sentences, slang words, and so on, much as a spelling or grammar checking program does. The formulas packaged with Microsoft Word for the Mac don't flag any of the aforementioned elements, however.

The Final Score

If we want our readers to get on base with our texts, we need to throw slow, straight pitches every time so that readers know what to expect. Readability formulas do not ensure that we are throwing straight, and in fact they can turn a good, straight pitch into a nasty curve ball. Revising a text to lower its reading grade level does not improve readability but can actually reduce it by making connected, flowing thoughts into a series of short, choppy sentences ("Dick hit the ball. Run, Dick, run!") So, if you are ever asked to adjust your text to meet a target reading grade level, don't. Be your own umpire and call 'em like you see 'em.

References

1. Ancker J. Developing the informed consent form: a review of the readability literature and an experiment. *AMWA J.* 2004;19:99-100.
2. Doak C, Doak L, Root J. *Teaching Patients with Low Literacy Skills.* Philadelphia, Pa: J.B. Lippincott, 1985.
3. Doak CC, Doak LG, Friedell GH, Meade CD. Improving comprehension for cancer patients with low literacy skills: strategies for clinicians. *CA Cancer J Clin.* 1998;48:151-162.
4. Redish JC, Selzer J. The place of readability formulas in technical communication. *Tech Comm.* 1985;32(4):46-52.
5. Lang T, Talerico C. Improving comprehension: theories and research findings. In: American Medical Writers Association. *Selected Workshops in Biomedical Communication, Vol. 2.* Bethesda, Md: American Medical Writers Association; 1997; 105-114.

A MYTH IS AS GOOD AS A MILE By Edie Schwager

In the movie "Casablanca," Rick Blaine (Humphrey Bogart) never said "Play it again, Sam." He said (rather petulantly, I thought) "Play it, Sam." The title of Woody Allen's movie is factually incorrect. (I gave a grammatical shudder when I saw the title of the Rick Moranis movie "Honey, I Shrunk the Kids.")

A. Conan Doyle wrote 4 novels and 56 short stories about Sherlock Holmes and his physician-friend, John H. Watson, but he never wrote "Elementary, my dear Watson." It was the incomparable Basil Rathbone as the quintessential sleuth who uttered those words in a movie. We Rathbone and S.H. buffs have to keep the record straight.

This information is taken from the preface to Richard Shenkman's book, *I Love Paul Revere, Whether He Rode or Not* (1991). His explanation about the title goes as follows:

> An iconoclast had noted [a few days before Harding's speech] that Revere never completed the ride made famous by Longfellow. Before giving warning to Concord, Revere was discovered by the British and captured. [Fast forward to the setting, Warren G. Harding's campaign stop in 1923.] Harding, however, told the crowd he didn't care. "I love the story of Paul Revere," the president intoned in his most presidential-sounding voice, "whether he rode or not."

Also from Shenkman's book (p. 29):

> Coca-Cola, because of the name and the secrecy surrounding its formula, was long suspected of containing cocaine. The Coca-Cola Company admitted as late as 1903 that its soda formula included cocaine, but the drug was used in exceedingly small amounts—so small that when federal inspectors conducted tests on the soda at the turn of the century they could not find even a trace of cocaine in the product.

"Something which everybody accepts as the gospel truth is inevitably false."

—H(enry) L(ouis) Mencken