

# DEVELOPING THE INFORMED CONSENT FORM: A REVIEW OF THE READABILITY LITERATURE AND AN EXPERIMENT

By Jessica Ancker, ELS, MPH

*Traumatic Brain Injury Clinical Trials Network  
Department of Biostatistics, Mailman School of Public Health, Columbia University*

Institutional Review Boards (IRBs) rightly insist that informed consent forms be easily readable by patients. IRBs generally assess readability with the Flesch-Kincaid Grade Level scale, which assigns an easy-to-interpret grade level from 0 to 12, corresponding to US school grades from kindergarten through high school.

However, usability and document design research suggests that a better way to assess the readability of a document is to study the people who read it. One method for studying readers is verbal protocol analysis, in which people are invited to read the document aloud, and their comments are analyzed to determine whether they understood the content and what they thought about it. Other methods are comprehension quizzes to determine whether readers can answer questions about the content, and usability testing to determine whether they can complete tasks described in the document. Any of these methods can be used in an iterative fashion to revise and improve a document or Web site.

As a writer for an academic research group, I was charged last year with writing an informed consent form for a multicenter study of patients with traumatic brain injury. My preference was to use document design research methods, but the form also had to meet an 8th-grade reading level as assessed by the Flesch-Kincaid scale. I ended up using a 2-step development process. In the first step, readers provided feedback on the form through verbal protocol analysis. In the second step, the draft was edited to the 8th-grade reading level.

This article describes the 2-step process and some of its results, including tips for getting a good Flesch-Kincaid grade. It begins with a discus-

sion of the development and limitations of readability scales.

## EARLY RESEARCH ON READING

The Flesch Reading Ease Scale, published in 1949, uses average sentence and word length to compute a score from 0 (“practically unreadable”) to 100 (very easy).<sup>1</sup> Rudolf Flesch derived the scale after studying reader performance on comprehension quizzes. Flesch described the scale as no more than a “simple yardstick,” arguing that really good writers should instead study their readers. “[M]ake it a habit to try your stuff on the cleaning woman or the elevator man,” he wrote.<sup>1</sup>

In the early 1970s, J. P. Kincaid adjusted the formula and normed it on US Navy recruits to produce the Flesch-Kincaid Grade Level scale.<sup>2</sup> The grade levels are calculated with this formula:

**Reading grade level = (0.39 x average sentence length) + (11.8 x average syllables per word) – 15.59**

The scale can produce grades of 16 or higher, indicating a college or graduate-school reading level. However, the version available in Microsoft Word truncates the scale at 12, lumping together all reading levels above high school.

Although Flesch hoped that readers “won’t take the formula too seriously and won’t expect from it more than a rough estimate,”<sup>1</sup> some in the medical community take readability formulas very seriously. For example, a recent article in *The New England Journal of Medicine* criticized consent forms solely because they had poor readability scores.<sup>3</sup> A systematic review of research on informed consent describes at least 17 similar articles.<sup>4</sup>

## NEWER CRITICISMS OF READABILITY FORMULAS

However, the nearly 30-year-old Flesch-Kincaid scale and other readability formulas are considered antiquated by reading researchers. In browsing through my university’s psychology library, I found no texts on the psychology of reading that cited Kincaid’s work or any other grade level scale.

Reading researchers have largely abandoned these mathematical formulas in favor of empirical research on readers.<sup>5,6</sup> One empirical research method is the analysis of verbal protocols, in which readers are recorded saying everything that they think as they decipher a document. Another empirical technique is usability testing, in which readers are observed as they attempt to perform the tasks described in a document. Schriver describes an excellent set of usability tests on a videocassette recorder (VCR) instruction manual: most users were unable to program the VCR until after the research team redesigned the manual.<sup>5</sup>

Here are some of the reasons why a good grade with a readability formula does not necessarily mean that a document is readable.

**Readability scales ignore the reader.** Reading is an active process of sense-making by a reader, but a text-based scale cannot measure anything about the reader. For example, such scales cannot detect whether the reader successfully programs the VCR.

The formulas also cannot assess characteristics that will affect how a reader makes sense of text, including his or her literacy level, fluency in English, familiarity with the topic, motivation, trust in the medical establishment, anxiety level, or emotional response to the text.<sup>5,7-9</sup> According to a readability scale, a sentence such as

“Be prepared to die next month” is easier to read than “Call for an appointment next month” simply because it contains shorter words.

**Readability scales equate word and sentence length with readability.** Readability scales depend on the assumption that shorter words are always more readable than longer words and that words of the same syllable count are equally readable. For example, “lesion” and “cancer” are assumed to be equally readable, even though “cancer” is more familiar to most audiences.<sup>7</sup>

**Readability scales ignore text organization.** The scales focus exclusively on words and sentences, ignoring larger issues that have more affect on readability: the document’s contents, sentence structure, paragraph structure, and organization.<sup>6,10</sup> The scales cannot determine whether the document contains logical gaps, provides adequate context for new ideas,<sup>9</sup> or meets readers’ cultural expectations.<sup>7</sup>

**Readability scales ignore visual design.** Readability scales say nothing about legibility or visual design elements, such as text size, font style, page layout, and use of illustrations, all of which are important to readers.<sup>5,6,10</sup> For example, the Flesch-Kincaid grade of this article doesn’t change if all the breaks between paragraphs are deleted. The scale also gives the same grade to these two passages:

#### Version 1

WE ARE INVITING YOU TO BE IN A RESEARCH STUDY BECAUSE YOU HAVE A TRAUMATIC BRAIN INJURY. A TRAUMATIC BRAIN INJURY IS AN INJURY TO THE BRAIN CAUSED BY SOMETHING THAT HIT OR SHOOK THE HEAD. THIS CONSENT FORM EXPLAINS THE PURPOSE, RISKS, AND BENEFITS OF THE STUDY. THIS INFORMATION MIGHT HELP YOU DECIDE WHETHER TO BE IN THE STUDY. PLEASE READ THIS FORM CAREFULLY. IF YOU HAVE ANY QUESTIONS, ASK YOUR DOCTOR BEFORE YOU MAKE A DECISION.

#### Version 2

We are inviting you to be in a research study because you have a *traumatic brain injury*. A *traumatic brain injury* is an injury to the brain caused by something that hit or shook the head.

This consent form explains the purpose, risks, and benefits of the study. This information might help you decide whether to be in the study.

Please read this form carefully. If you have any questions, ask your doctor before you make a decision.

**Readability formulas are better for assessing writing than for guiding revisions.** Early researchers applied the scales to existing texts, such as school textbooks, and found that children of a certain grade performed best on comprehension quizzes when text had a certain average word length and sentence length.<sup>2</sup> Researchers sometimes found that the converse was also true: that is, when short excerpts from age-appropriate texts were revised to meet word-length and sentence-length specifications, children in the target grade performed better on comprehension tests.<sup>11</sup>

However, literature reviews published as early as 1985 documented that readability formulas are unreliable and do not predict the difficulty of full-length documents read by adults in real-world situations.<sup>10</sup> Even more important, lowering a document’s grade level without making any other changes to it produces little to no improvement in comprehension.<sup>7,8,10,12,13</sup>

By contrast, comprehension of informed consent forms can be improved by combining grade-level changes with visual redesign, changes in sentence structure, revisions directed by focus group testing, or the addition of illustrations.<sup>14,15</sup> It is not clear how much, if any, of the improvement is attributable to the grade-level reduction.

In brief, shortening the words and sentences alone cannot salvage a document that discusses complex concepts with insufficient context and background, uses unfamiliar words, is poorly organized, or contains insufficient information.

## A PROCEDURE FOR DEVELOPING A READABLE INFORMED CONSENT FORM

Nevertheless, in developing an informed consent form, I had to use the Flesch-Kincaid scale because some of the hospitals participating in our multicenter study have IRBs that require it. I decided to develop the form in a 2-step process, first soliciting reader feedback through protocol analysis and only then applying the Flesch-Kincaid scale. The first draft was a 2.5-page version adapted from a consent form from a similar study, formatted in single-spaced 12-point Times Roman with bold-faced subheads.

### Step 1: Using protocol analysis.

I had drawn on my own experience to decide how to create an approachable and understandable document, but readability research has repeatedly shown that readers make sense of documents in ways that are not anticipated by writers.<sup>5</sup> Therefore, in the protocol analysis experiment, I invited 3 people to read the form aloud to me. I noted their questions, comments, hesitations, and mistaken interpretations. After revising the draft, I gave a second version to 2 additional readers and then revised it again.

Two of the readers had attended college, 3 had not. Two were native speakers of English, 1 used English as a second language, and 2 were bilingual in Spanish and English. The Columbia Presbyterian Medical Center IRB approved this research project and declared it exempt from IRB supervision.

Following are some examples of changes prompted by the readers.

- **Organization:** The first draft described the study goals near the beginning of the document, but most of the readers immediately asked what the study would require of them. I moved the section “What will happen if I agree to be in the study?” closer to the beginning.

- **Language:** The reader who spoke English as a second language stumbled over the phrase “This study poses no risks.” Although she did not ask any questions, her hesitation indicated that “pose” was unfamiliar. I reworded the phrase to read, “This study has no risks.”
- **Emotional tone:** The readers frequently commented on their emotional reactions to the form, and some shared anecdotes about unpleasant interactions with the medical system or medical researchers. Two said that the paragraphs about risks and confidentiality made them feel anxious and mistrustful. Anxiety could be a good thing if it indicated that readers were being prompted to think seriously about the warnings. However, mistrust suggests that they were suspicious of the researchers’ motives. I reorganized the document so that readers encountered the warnings only after reading assurances that they were free to drop out of the study at any time. The reorganization may have given readers more of a sense of control, because the second group of readers did not raise the issue. I considered it ethical to make the change because the study carries no health risks and poses little risk to confidentiality. However, for studies with substantial risks to health or confidentiality, downplaying risks would be unethical.

## Step 2: Applying the Flesch-Kincaid scale.

The resulting form earned a Flesch-Kincaid grade of 8.7. To be acceptable to all the IRBs, it needed a grade of 8.0.

I applied the automated Flesch-Kincaid scale in Microsoft Word. (To display the Flesch-Kincaid grade, click on the “Spelling and Grammar” tool, which calls up the spell-check dialog window. Click the “Options” button at the bottom of the dialog window, then

check the “Show readability statistics” box. Click “OK” and continue with the spell-check. When the spell-check is finished, a “Readability Statistics” window will appear with the results.)

The Flesch-Kincaid scale was extremely sensitive to small changes in the draft, which could prompt some silly decisions. For example, replacing all 11 instances of the 4-syllable word “information” with the 2-syllable word “data” lowered the document’s score by almost a full grade. However, in this case, the shorter “data” would be a poor choice because it is much less familiar than the longer “information.” (Consider which word you would use when talking to a second-grader!) “Data” also conveys an awkward, formal feeling. For example, the global change produced this passage: “This consent form explains the purpose, risks, and benefits of the study. These data might help you decide whether to be in the study.”

Also, a version of the consent form that explained the term “traumatic brain injury” scored worse than a version that provided no explanation. This finding might encourage writers to omit explanations of complicated medical terms, which clearly is not a good idea. In fact, several of the pilot readers said they liked having a definition in the document (“an injury to the brain caused by something that hit or shook the head”).

The grade level could be brought even lower by dropping the term “traumatic brain injury” altogether. Medical terms are often multisyllabic, and omitting them will lower the Flesch-Kincaid grade. This, too, would be a bad idea, because medical terms will be part of the patient-physician discussion that is integral to the informed consent process. If the document and the physician use different vocabulary, the patient could end up confused.

**Beating the system by cutting up sentences:** After experimentation, it became clear that the best way to change the Flesch-Kincaid grade was to break up a few long sentences and

add a few short sentences and sentence fragments. Flesch himself told writers that they could get good scores only by varying sentence length.<sup>2</sup>

The Flesch-Kincaid scale counts subheadings as short sentences. The document got a lower grade when I inserted subheads, such as “What are the risks and benefits of the study?” Fortunately, subheads are good for the reader because they introduce topics and help the reader locate information.

Breaking up long sentences into lists also reduces the Flesch-Kincaid grade. This description of the study goals gets a 12.0 because it is long, even though it is structurally very simple:

“We want to know how many people suffer these injuries at different times of year, what types of brain injuries they suffer, how long it takes them to reach the hospital, and how long they stay in the hospital.”

If the sentence is turned into a bulleted list, the Flesch-Kincaid scale counts each fragment as a short sentence and awards the list a grade of 2.3.

“We want to know

- how many people suffer these injuries at different times of year,
- what types of brain injuries they suffer,
- how long it takes them to reach the hospital, and
- how long they stay in the hospital.”

However, as noted earlier, most of the pilot readers said that they were more interested in the study’s effect on them than in the researchers’ goals. I decided that the study goals paragraph did not warrant the extra emphasis of a bulleted list.

**The final cut:** At one point, the document had a grade of 8.1. I scanned through it and found this long sentence: “Your name and your medical records number will be kept in locked drawers at your hospital, and only research staff will be able to see them.” Breaking the sentence into two sentences, at the comma, brought the

grade of the entire document from 8.1 down to 8.0. What a disproportionate impact for such a small change!

**IRB decision:** In the end, the IRBs decided that the study did not require a full-length informed consent form, although one required patients to sign an abbreviated form acknowledging that they had received an oral description of the study. Thus, unfortunately, no information is available about the real-world usefulness of the form described in this article. Soliciting feedback from actual patients in the research setting would be the best way to continue to improve an informed consent document.

## CONCLUSION

Working with the Flesch-Kincaid scale vividly demonstrated its limitations. The scale cannot predict readers' reactions or account for the quality or accuracy of the information, the document's visual design, or its cultural and rhetorical appropriateness. The scale is sensitive to word length, but it cannot identify short words that might be unfamiliar. The scale is even more sensitive to sentence length, so that dividing a single sentence can change the grade of an entire document.

In my opinion, useful information about the success or failure of the document came from pilot testing it with readers who were similar to the intended audience, not from the Flesch-Kincaid grade. Document and Web site usability testing is based on the premise that users are the best source of information about usability.<sup>5</sup>

Flesch-Kincaid grades may be somewhat useful in the research world because they might alert researchers to convoluted sentences and complicated words. Poorly written documents often get high grades. Unfortunately, if inexperienced writers revise their documents to get a better grade, they may make inappropriate decisions, such as omitting important medical terms, skipping explanations, or replacing familiar words with shorter but less familiar ones. Using the Flesch-Kincaid

scale might also encourage writers to ignore the document's overall design and organization, because such factors do not affect the grade.<sup>10</sup> The scale is no substitute for empirical tests with real readers, whose interpretations will frequently surprise even the most experienced writer.

Medical writers will probably benefit from playing around with the Flesch-Kincaid scale to find out how it works but should take care not to violate proven principles of readability just to get a good grade. Clear, simple, understandable writing may get a good grade, but focusing on the grade won't result in good writing.

*This article has a grade of 11.0 on the Flesch-Kincaid scale. However, I got more useful feedback about its readability by pilot testing it on my husband, John Affleck. Thanks, John!*

## Acknowledgments

This work was supported by the Traumatic Brain Injury Clinical Trials Network of the National Institute of Child Health and Human Development. Many of the ideas in this article were developed in discussions with Tom Lang. Thanks are also due to an anonymous peer reviewer, whose comments were thorough and thoughtful.

## References

1. Flesch R. *The Art of Readable Writing*. 25th Anniversary Edition. New York: Harper & Row; 1974.
2. Kincaid JP, Fishburne RP, Rogers RL, Chissom BS. Derivation of new readability formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy enlisted personnel. Research Branch Report 8-75. Memphis: Naval Air Station; 1975.
3. Paasche-Orlow MK, Taylor HA, Brancati FL. Readability standards for informed-consent forms as compared with actual readability. *N Engl J Med*. 2003;348:721-726.
4. Sugarman J, McCrory DC, Powell D, et al. Empirical research on informed consent: an annotated bibliography. *Hastings Cent Rep*. 1999;29:S1-S42.
5. Schriver KA. *Dynamics in Document Design*. New York: John Wiley & Sons; 1997.
6. 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.
7. 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.
8. Taub HA, Baker MT, Sturr JF. Informed consent for research. Effects of readability, patient age, and education. *J Am Geriatrics Soc*. 1986;34:601-606.
9. Gopen GD, Swan JA. The science of scientific writing. *Am Sci*. 1990;78:550-558.
10. Redish JC, Selzer J. The place of readability formulas in technical communication. *Tech Comm*. 1985;32:46-52.
11. Flesch R. *How to Test Readability*. New York: Harper & Bros.; 1951.
12. Davis TC, Holcombe RF, Berkel HJ, et al. Informed consent for clinical trials: a comparative study of standard versus simplified forms. *J Natl Cancer Inst*. 1998;90(9):668-678.
13. Young DR, Hooker DT, Freeburg FE. Informed consent documents: increasing comprehension by reducing reading level. *IRB* 1992;12:1-5.
14. Murphy DA, O'Keefe ZH, Kaufman AH. Improving comprehension and recall of information for an HIV vaccine trial among women at risk for HIV: reading level simplification and inclusion of pictures to illustrate key concepts. *AIDS Ed Concepts*. 1999;11:389-399.
15. Dresden GM, Levitt MA. Modifying a standard industry clinical trial consent form improves patient information retention as part of the informed consent process. *Acad Emerg Med*. 2001;8:246-252.