

## Assessing Readability of Consumer Health Information: An Exploratory Study

Darren Gemoets, Graciela Roseblat, Tony Tse, Robert Logan

*Lister Hill National Center for Biomedical Communications, National Library of Medicine, Bethesda, MD, USA*

### Abstract

*Researchers and practitioners frequently use readability formulas to predict the suitability of health-related texts for consumers (e.g., patient instructions, informed consent documents). However, the appropriateness of using readability formulas—originally developed for students and educational texts—for lay audiences and health-related texts remains to be validated. In this exploratory study, we compared two methods of assessing the readability of consumer health materials: the Cloze procedure, using actual readers, and readability formulas, using our Readability Analyzer program. A statistically significant inverse correlation ( $r = -0.581$ ,  $p = 0.01$ ) was found, suggesting that the Readability Analyzer may provide a reasonable “first approximation” for predicting readability of consumer health texts. We also identified several linguistic factors associated with increased reading ease as candidates for improving the performance of the Readability Analyzer. Our ultimate objective is to develop tools to support the design and evaluation of health information that is comprehensible and accessible to laypersons.*

### Keywords:

Readability, Patient Education, Linguistics, Evaluation Studies, Communication Barriers.

### Introduction

Consumers are increasingly using online health information systems where most content is textual. A recent Pew survey [1] estimated that approximately 77 million adults in the U.S. have sought health-related information online as of March 2003. Thus, in addition to ensuring quality content (e.g., accuracy and currency), information providers must consider whether laypersons will understand such texts, which often include complex medical terms and concepts. Ideally, all consumer health information should consider factors such as the audience's domain knowledge, reading ability, and socio-cultural background. Until this is possible, authors need to adjust their writing styles, and intermediaries (e.g., healthcare professionals, medical librarians) must take care to select documents that consumers can understand.

Educational researchers have developed methods of assessing readability or the difficulty of text to determine the appropriate grade level for students [2]. In other words, educators desire the best match between texts and the reading abilities of students at a particular grade level. *Readability formulas*, simple algorithms designed to help educators objectively determine the

grade level of texts, are frequently used to assess readability in many domains, including biomedicine. For example, a PubMed search on “readability” (3 Sept 2003) retrieved 1,585 citations, from such diverse disciplines as oncology nursing, medical legal issues, and communication disorders. A recent systematic review of studies on the quality of consumer health Web sites identified several distinct studies that used readability formulas explicitly as part of the evaluation methodology [3]. As Doak, Doak, and Root [4] point out, “Readability formulas offer the health care provider an easy-to-use method to assess the reading difficulty of most print materials.” Furthermore, these formulas are readily available through office-productivity applications and the Web.

While we focus on the difficulty of English text, interest in readability is international. For example, the Fry Readability Graph has been validated in Spanish. Berland, et al. [5] compared readability of both English and Spanish online sources. The Canadian Public Health Association's Plain Language Service (PLS<sup>1</sup>) offers a variety of services in English and French from document assessment (including readability via the SMOG readability formula) to workshops on plain language and verbal communication techniques. The DISCERN Project at the University of Oxford<sup>2</sup> provides a tool that helps to evaluate the quality of a publication's content, and has been translated into German.

Despite the popularity of readability formulas for assessing consumer health materials, the suitability of these tools in this domain has been an unvalidated assumption. Traditional formulas concentrate on factors such as words per sentence and syllables per word (or proportion of “difficult” words). Concerns have been raised about the effectiveness of these formulas in general—and for consumer health texts in particular. For example, other factors may affect reading comprehension, such as style (e.g., active and passive voice, first and third person), linguistic structure (e.g., text coherence, general discourse structure), and extratextual features (e.g., bulleted lists, font size). In health-related texts, domain-specific issues include the reader's medical knowledge, personal experience, and understanding of technical terms, acronyms, and semantic relationships [3,6]. Even psychological factors may impact comprehension, such as reader motivation and emotional reaction to content, presentation, or source.

- 
1. The PLS is a service of the National Literacy and Health Program (<http://www.nlhp.cpha.ca/>).
  2. See DISCERN's web site at <http://discern.dphpc.ox.ac.uk/>.

In this exploratory study, we evaluate the commonly accepted practice of applying scores from readability formulas to assess consumer health materials. Specifically, we:

- Validated the use of readability formulas to assess reading difficulty of consumer health texts against the Cloze procedure;
- Investigated the effects of four consumer health genres, drawn from the field of mass communication; and
- Identified linguistic factors that may improve readability assessment of consumer health materials.

Our long-term goal is to incorporate substantial factors that predict the readability of consumer health texts, including those specific to the medical domain, into our in-house Readability Analyzer tool.

## Background

### *The Cloze procedure and readability formulas*

Taylor [7] introduced the Cloze procedure to assess readability. In this procedure, representatives of an intended audience are given text passages with words removed at periodic intervals (e.g., every fifth word) and asked to fill in the blanks with context-appropriate words. Readers at the appropriate levels are able to provide “closure” to the passage based on information redundancy inherent in communication. That is, readers who are well-matched with the passage are able to construct sufficient meaning from the remaining text through various syntactic, semantic, cultural, and pragmatic clues to supply the omitted words [8]. As noted in [4], “The ability of readers to fill in missing words correctly is a valid indicator of how well they understand the passage (p. 35).”

Researchers in various domains have validated Cloze and readability formulas against each other. In particular, the Cloze procedure was originally validated against the Fry and Dale-Chall readability formulas using literary passages [7]. Conversely, readability formulas have been validated against the Cloze procedure, mainly in the education domain [2]. However, to our knowledge, no such validation has been conducted in the consumer health domain.

Five readability formulas were selected for this study: the revised Fry Readability Graph, Flesch Reading Ease, Flesch-Kincaid, Gunning FOG, and New Dale-Chall formula. These formulas represent various strategies for grade level determination and for computing semantic difficulty of text— including syllables per word, number of words with three or more syllables, and proportion of words appearing on a fourth-grade vocabulary list. Moreover, these formulas are well-known, easy to use, and designed for reading passages for primary and secondary school students.

The Readability Analyzer (RA), which implements a number of readability formulas, is an in-house tool written in Java with a Web-based front end. This tool uses tokenization and variant-generation software developed at the National Library of Medicine and publicly available syllable-counters<sup>1</sup>. To calculate a passage’s overall grade level, the RA simply averages grade levels across formulas, excluding any that are outside a formula’s designed range.

## Genre

Different types of health materials target particular lay audiences. In the mass communication literature, these genres are implemented through different writing and editing styles [9] and are differentiated by interactions among several factors:

- Accessibility to a mass medium;
- Intended audience; and
- Author’s narrative style.

For example, while newspapers are less expensive and more widely distributed than books, they must appeal to a more diverse audience with different levels of interest, domain knowledge, and literacy. On the other hand, books may appeal to a subset of readers (e.g. patients and caregivers) who may be more attentive to specific health topics. For lay audiences, the most accessible narrative style tends to include less background, convey the most important information first (e.g., wire service news style), and introduce a case study to symbolize the topic and engage the reader (e.g., news feature story) [10].

As no standard definitions were found in the literature, we operationalized four consumer health genres with the following criteria:

- **Public Broad**  
Health materials that are readily available, inexpensive, widely disseminated, and intended for a general circulation audience. The style conveys the topic’s newsworthiness to general audiences. Examples include health-related wire service news, and news or feature stories, both in print and online.
- **Public Limited**  
Health materials that are readily accessible, but less available to a general circulation audience. While sharing stylistic elements of news and feature writing, initial sentences are written to draw attention to a product or an organization. Examples include press releases about health topics for members of a target audience, such as investors and journalists.
- **Patient Broad**  
Health materials intended for patients and caregivers, but easily available to consumers. The style excludes news or feature stylistic characteristics and does not contain extensive background or context to interpret medical instructions, often formatted as a list. Examples include instructions, fact sheets, and newsletters for patients and caregivers that are written for a lay audience and freely distributed by voluntary health organizations and health-care networks.
- **Patient Limited**  
Health materials intended exclusively for patients and caregivers, and thus less available, requiring more effort to obtain. The style includes instructional information, extensive background materials, and technical jargon.

1. MMTx (<http://mmtx.nlm.nih.gov/>) for tokenization; LVG (<http://umlslex.nlm.nih.gov/lvg/>) for variants; University of Arizona word list and Lingua-EN-Syllable-0.251 for syllable counting.

Examples include self-help guides, reference books, and health newsletters.

### **Linguistic factors of readability**

Clustering documents by consumer health genre provides a convenient way to assess stylistic and linguistic factors. Other research has examined linguistic characteristics of different medical genres, including journal articles, review articles, textbooks, and “patient information” [11,12]. Their results suggest that many linguistic factors may impact readability, such as sentence and word length, prepositional phrases, personal pronouns, relative clauses (clauses that contain *who*, *that* or *which*), passive voice verbs, and lexical density<sup>1</sup>.

We focus on those linguistic factors that impose the least demands on human information processing, promoting comprehension.

## **Methods**

### **Document selection**

Twenty consumer health documents were selected from mass media print and online sources such as newspapers, magazines, reference/self-help books, and pamphlets. Selection was based on topic and genre (Table 1). Two topics—*allergies* (common condition) and *celiac disease* (less common)—control for familiarity or prior knowledge of the topic among participants. The four consumer health genres, characterized by diverse styles, represent text for different intended audiences and document availability. The operational definitions of the four genres were pre-tested and the criteria refined. For the study, the investigators selected exemplars of each of the four genres and avoided writing samples that could potentially be classified within more than one genre.

Table 1: Documents selected by topic and genre

| Genre                            | Allergies | Celiac Disease |
|----------------------------------|-----------|----------------|
| Public Broad (“feature story”)   | 3         | 2              |
| Public Limited (“press release”) | 3         | 2              |
| Patient Broad (“health column”)  | 3         | 2              |
| Patient Limited (“home guide”)   | 3         | 2              |
| Total                            | 12        | 8              |

### **Passage generation**

To prepare the 20 documents for assessment by the Cloze procedure and Readability Analyzer, and to mask document source and genre, all titles, headings, subheadings, and other metadata (e.g., author names) were removed. Raw text was extracted from its surrounding format and presentation, except for bullets. Twenty text passages, consisting of the first 250 words of text, or the closest number of words without breaking sentences, were created from the 20 documents. Enough text remains to retain the style and linguistic factors that characterize each genre, without overly taxing participants. The 250-word cut-off is also standard for the Cloze procedure.

### **Data collection**

While the Readability Analyzer automated the collection of average grade level scores for each text, Cloze procedure results had to be collected manually. A convenience sample of 40 participants was recruited from the National Library of Medicine, excluding medical experts. Each participant completed two Cloze passages. Following standard Cloze procedure, we only counted exact participant-author word matches and calculated Cloze scores as shown in Equation (1).

$$\text{Cloze score} = (\text{exact matches} / \text{blanks}) \times 100 \quad (1)$$

By counting only exact matches, we eliminated bias in our scoring of the Cloze passages. To facilitate human scoring, we wrote a Java program to generate the excerpts and corresponding keys for all 20 Cloze passages. To evaluate document validity and test duration, three members of the research team performed the Cloze procedure on all passages before distributing them to participants. This led to the substitution of one text that lacked sufficient context and contained numerous proper names and irrelevant details.

### **Data analysis**

Differences in scores from the Cloze procedure and Readability Analyzer, across all 20 passages and specifically among genres, were analyzed for statistical significance with Microsoft Excel and SPSS, a statistical software package.

For each passage, we performed raw frequency counts of linguistic factors reported in the literature to affect reading comprehension [11,12]. The results were compiled and considered alongside the results of the comparison between the Cloze procedure and Readability Analyzer.

## **Results**

### **Overall comparison of Cloze procedure and Readability Analyzer**

Despite a high variance in Cloze scores (SD = 11.7), the Cloze and Readability Analyzer (RA) scores across all 20 consumer health passages were statistically significantly inversely correlated ( $r = -0.581$ ,  $p = 0.01$ , two-tailed), as shown in Figure 1. As comprehension increases (measured by higher Cloze scores), RA grade level scores decrease, indicating more readable text.

### **Comparison of Cloze procedure and Readability Analyzer between genres**

In all but one pair-wise comparison, the differences between the Cloze and RA scores for the genres were found to be statistically significantly different ( $p < 0.05$ ). Mean scores by genre are summarized in Table 2.

Table 2: Summary of readability scores by genre

| Genre           | Cloze (%) | RA Score (Grade) |
|-----------------|-----------|------------------|
| Patient Broad   | 57.4      | 9.9              |
| Public Broad    | 49.4      | 12.5             |
| Public Limited  | 43.7      | 13.5             |
| Patient Limited | 39.5      | 16.5             |

1. Lexical density measures the proportion of content-loaded words in text, which contributes to text complexity [13].

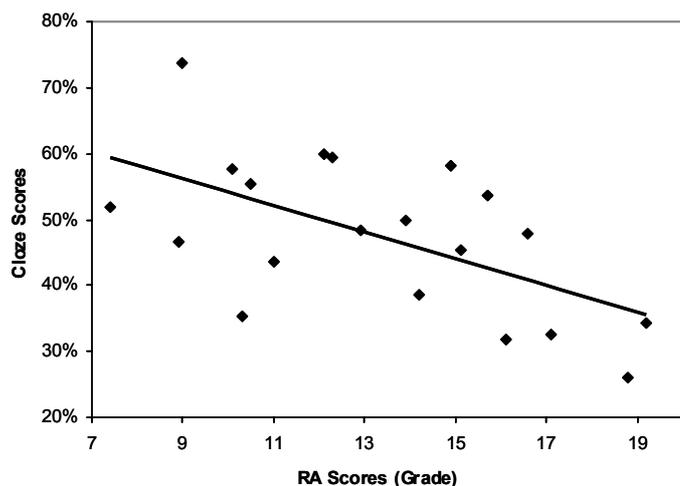


Figure 1 - Correlation between Cloze and RA scores across 20 consumer health passages

### Linguistic factors clustered by genre and readability

The 11 linguistic factors observed to have the greatest effect on readability by genre and their values are listed in Table 3. With the exception of Syllables/Word and Lexical Density, all values in the table represent average occurrences per document, broken down by genre.

Table 3: Frequency of linguistic factors by genre

| Genre                         | Patient-Broad | PublicBroad | Public Limited | Patient Limited |
|-------------------------------|---------------|-------------|----------------|-----------------|
| Epistemic Modals <sup>a</sup> | 4.4           | 1.8         | 1.2            | 2.6             |
| Personal Pronouns             | 7.5           | 4.4         | 3.2            | 0.6             |
| Proper Nouns                  | 0.2           | 11.8        | 7.2            | 2.2             |
| Words / Sentence              | 14.5          | 21.5        | 19.9           | 18.8            |
| Prep. Phrases                 | 22.3          | 28.2        | 27.8           | 28.8            |
| Relative Clauses              | 2.4           | 3.0         | 3.0            | 4.2             |
| Passive Voice                 | 0.8           | 1.8         | 1.0            | 1.8             |
| Use of Statistics             | 0.6           | 1.0         | 1.0            | 1.4             |
| If-Then Props.                | 1.4           | 0.4         | 1.2            | 0.2             |
| Syllables / Word              | 1.58          | 1.63        | 1.77           | 1.97            |
| Lexical Density               | 44.4%         | 47.9%       | 53.2%          | 54.1%           |
| RA Grade Level                | 9.9           | 12.5        | 13.5           | 16.5            |

a. E.g., *may*, *could*, and *can*.

## Discussion

Scores averaged from five readability formulas, as implemented in the Readability Analyzer, were statistically significantly inversely correlated with Cloze procedure scores across the full range of documents, as well as between genres. Because our study participants were mostly professionals, this finding may not generalize to a typical consumer population. Pending validation on a larger scale, and with actual consumers for whom these materials are intended, this finding suggests that RA scores are reasonable surrogates for Cloze procedure scores for text passages extracted from consumer health documents. To our knowledge, no other research has attempted to validate readability formulas with the Cloze procedure for consumer health-related

materials. Two implications stem from these preliminary findings:

1. Readability formulas appear to serve as a reasonable first approximation for predicting how well consumers will understand textual materials about health topics<sup>1</sup>.
2. Readability formulas may replace the resource-intensive Cloze procedure, which requires preparation of text passages and availability of participants.

Writers in each genre have different purposes and intended audiences, favoring linguistic devices that best serve their communicative needs. Our linguistic analysis suggests that clusters of features may differentiate the four genres used in this study, especially between the genre with the best readability (i.e., “easiest”) and the other three. In our corpus, texts in the Patient Broad genre obtained the best readability scores as measured by both the Cloze procedure and Readability Analyzer (Table 2). These texts exhibit many linguistic features that maximize comprehension and minimize cognitive load. Specifically, these texts have the following characteristics:

- Shortest average **word** and **sentence length**;
- Most **epistemic modals**<sup>2</sup>;
- Lowest **lexical density** [13], fewest **relative clauses** [14] and fewest **prepositional phrases**, all of which correlate significantly with better reading comprehension and lower text difficulty;
- Lowest frequency of text-embedded **statistics**;
- Most **if-then propositions**: the simplest way to express causal relationships, allowing for their immediate recognition;
- Most **personal pronouns** and fewest **proper nouns**, indicative of the author's more personal tone and direct dialogue with the reader; and
- Fewest **passive voice** sentences, which are wordier than their active counterparts and more difficult to process due to the reversal of the usual Subject-Verb-Object word order in English [15].

Based on our preliminary finding that readability formulas are a viable empirical tool to measure readability of consumer health texts, the Readability Analyzer could be enhanced to account for additional linguistic and cognitive factors. These additional factors could be weighted with syntactic and semantic factors from traditional readability research to provide a more comprehensive and accurate evaluation of consumer health materials.

## Conclusion

The statistically significant inverse correlation between Readability Analyzer and Cloze procedure scores suggests external validation of traditional readability formulas in the consumer

1. Of course, readability should be only one of a number of considerations when assessing consumer health materials.
2. Although epistemic markers have long been associated with hedging devices—author's expression of certainty or likelihood about a proposition [11,12]—the modals in these texts introduce symptoms/signs that the readers may or may not experience.

health domain. Genres derived from mass communication helped to isolate linguistic factors most associated with readability in our pilot corpus. With a flexible tool like the Readability Analyzer to implement those and other factors of readability, we are poised to research each factor's relative impact on readability, weight each factor accordingly, and ultimately improve our automated assessment of the difficulty of health-related text to consumers.

## Acknowledgments

We would like to thank our 40 participants for their time to take part in our study. We would also like to thank Lisa Russell-Pinson and Olivier Bodenreider for their support and insights.

## References

- [1] Fox S, Fallows D. Internet health resources. Health searches and email have become more commonplace, but there is room for improvement in searches and overall Internet access. Washington, D.C.: Pew Internet & American Life Project; 2003.
- [2] Chall JS, Dale E. *Readability revisited: the new Dale-Chall readability formula*. Cambridge (MA): Brookline Books; 1995.
- [3] Eysenbach G, Powell J, Kuss O, and Sa ER. Empirical studies assessing the quality of health information for consumers on the World Wide Web: a systematic review. *JAMA*. 2002 May 22-29;287(20):2691-700.
- [4] Doak CC, Doak LG, and Root JH. *Teaching patients with low literacy skills*. 2nd ed. Philadelphia: Lippincott Williams & Wilkins; 1996.
- [5] Berland GK et al. Health information on the Internet: accessibility, quality, and readability in English and Spanish. *JAMA*. 2001;285(20): 2612-21.
- [6] Roberts JC, Fletcher RH, and Fletcher SW. Effects of peer review and editing on the readability of articles published in *Annals of Internal Medicine*. *JAMA*. 1994;272:119-121.
- [7] Taylor WL. Recent developments in the use of "cloze procedure." *Journal Q*. 1958;42-8.
- [8] Dale NB. A pilot study using the Cloze technique for constructing test questions. *Journal of Computers in Mathematics and Science Teaching*. 1999;18(3): 317-30.
- [9] McQuail D. *McQuail's mass communication theory*. 4th ed. Thousand Oaks (CA): Sage; 2000.
- [10] Moen D, Brooks BS, Kennedy G, Ranly D. *News reporting and writing*. New York: Bedford, St. Martins; 2001.
- [11] Biesenbach-Lucas S. *A comparative genre analysis: the research article and its popularization*. Unpublished Ph.D. dissertation. Washington, DC: Georgetown University; 1995.
- [12] Russell-Pinson L. *Linguistic and extratextual variation in medical English writings: a comparative genre analysis*. Unpublished Ph.D. dissertation. Washington, DC: Georgetown University; 2002.
- [13] Jenkins JR, Dixon R. Vocabulary learning. *Contemporary Educational Psychology*. 1983;18: 237-60.
- [14] Craghead NA, Donnelly KG. Comprehension of superordinate and subordinate information by good and poor readers. *Language, Speech, and Hearing Services in the Schools*. 1982;13:177-186.
- [15] Garner BA. *A dictionary of modern American usage*. New York: Oxford UP; 1998.

## Address for correspondence

Email: [gemoets@nlm.nih.gov](mailto:gemoets@nlm.nih.gov)