Original Paper

Text Features and Readability: Expert Evaluation of Consumer Health Text

Graciela Rosemblat, PhD; Rob Logan, PhD; Tony Tse, PhD; Laurel Graham, MS
National Library of Medicine, National Institutes of Health, Bethesda, MD, USA

Corresponding author: Graciela Rosemblat, PhD, grosemblat@mail.nih.gov

Abstract

Background: Previous research suggests that consumers frequently have difficulty understanding written health information.

Objectives: This exploratory study investigated the influence of linguistic and stylistic features on the readability of consumer health texts. Specifically, the research goals were (1) to examine the importance of previously identified predictors of general readability in the consumer health domain, based on expert judgment, and (2) to characterize patterns associated with expert ratings of readability across the various predictors.

Methods: Health communication experts (n = 4) reviewed a sample of 22 consumer health texts on two common health topics, asthma and weight management. Each expert independently rated the contribution of 15 specific features on the readability of all documents in the sample.

Results: Simultaneous multiple regression found that the 15-variable model significantly predicted readability for a general audience ($F(15, 72) = 11.802$; $P < .001$). Two variables, "Vocabulary" and "Main Point" significantly predicted general audience readability. A factor analysis of all ratings for the 15 features across the 22 documents revealed three clusters of features representing expert perceptual orientations: (1) discourse-level features, (2) sentential-level features, and (3) semantic features ("Vocabulary" and "Main Point").

Conclusions: The preliminary results suggest that developing consumer health-specific readability tools may require both modifying existing general measures, such as including health-related vocabulary, as well as adding new predictive features, such as ability to detect the "take-home" message. Future work includes verification of this expert evaluation by consumers.

Keywords**:** Comprehension; health education; Internet; evaluation studies; consumer health text

Introduction

Recent studies show that nearly two-thirds of all online English-speaking adults in the United States have reported seeking health information [1]. While consumers turn to the Internet for health information, nearly half of all American adults—90 million people—have difficulty understanding and using health information. In fact, more than 300 studies indicate that health-related materials cannot be understood by most of the intended audience [2].

General readability formulas (eg, Flesch-Kincaid, Dale-Chall) were developed to assess the grade level of general educational materials up to secondary school, and account for two textual features: syntax (eg, average words per sentence) and semantics (eg, familiar vocabulary terms) [3]. In a previous study, the authors examined the applicability of such general formulas to text in the consumer health domain. The results suggested that, while existing measures provide a reasonable "first approximation," other textual features such as word repetition may improve performance of readability formulas with consumer health text [4]. While many of these features

were explored in earlier readability studies [5], they have not been examined explicitly for health text written for consumers [8]. (See [5] for a review of readability tools in the health domain.)

The current study looks at how experts assess the relative contribution of 15 linguistic and stylistic factors identified in the literature (eg, [6,7] among others) to the readability of consumer health text (Table 1). Two research questions were posed:

1. What is the importance of previously identified predictors of general readability in the consumer health domain, based on expert judgment?

2. What patterns characterize the expert ratings of readability across the various predictors?

Table 1. Brief description of the 15 textual features evaluated in this study

Linguistic/Stylistic Feature	**Brief Description/Comment**
1. Vocabulary	Number of words that are likely to be familiar to readers
2. Lexical Density	Ratio of nouns and verbs to articles and prepositions
3. Word Repetition	Number of times a reader is exposed to a word
4. Prior Knowledge	Background knowledge needed to understand a text
5. Words per Sentence	Number of words in the text divided by the total number of sentences
6. Words per Package	Number of commas and parentheticals to parse complex sentences
7. Sentence Structure	Number of sentences with subject-verb-object order
8. Cognitive Load	Number of ideas within a sentence
9. Discourse Markers	Logical transition between and among sentences
10. Personal Pronouns	(as applicable)
11. Voice	Passive versus active voice
12. Bullets/Lists	(as applicable)
13. Concept Repetition	Repetition used to reinforce key concepts
14. Flow	Logical ordering of individual ideas to build a complex network of concepts
15. Main Point	Ability of readers to identify and understand the “take home” message

Methods

A convenience sample of 22 consumer health documents were selected from health insurance / healthcare organization Web sites. The texts, in two styles (narrative and instructional) and on two health topics (asthma and weight management), were selected to be representative of patient education materials available at these sites. The passages, ranging between the 4^{th} and 14^{th} grade levels, based on an averaged score among five general readability formulas [4] were normalized to approximately 400 words in length. Graphics, typography (font type, size), formatting elements (except for bullets, titles, headings and subheads), and other metadata (eg, author names) were removed to minimize potential confounding factors and mask information sources. The final text was presented in plain ASCII.

A total of four experts in health communication or “annotators” were contracted to evaluate the 15 pre-specified linguistic and stylistic features (ie, outcome variables) as measures of text readability across the 22 normalized consumer health passages. All annotators held doctorates in mass or health communication, had experience in health communication research (including

knowledge of the readability literature), and held academic positions. The reviews were conducted in February-April 2005. The annotators assessed:

- Independent Variables (for each 15 textual features): “How successful or unsuccessful are each of the following factors implemented in each article in fostering a high readability for general audiences, operationally defined as high school and middle school students?”
- Dependent Variable: “In your opinion, how successful or unsuccessful was the author in making this article readable for general audiences (eg, high school/middle school students)?”

All outcome variables were measured by five-point Likert scales (interval data) with responses ranging from “Very Unsuccessful” (1) to “Very Successful” (5). The first research question (features that predict readability) was assessed via a multiple regression model. Factor analysis was used to evaluate the second research question (patterns across all features). The Statistical Package for the Social Sciences (SPSS) was used to analyze all results.

Results

For the first research question, the authors first examined the regression data for normality and multicollinearity. No significant multicollinearity was identified. Further, no features (or independent variables) were skewed and no outliers exceeding three standard deviations were detected. Linear regression was used in this study because the data were normally distributed and interval. The authors entered the 15 independent variables into the regression model simultaneously (enter method), but given the exploratory nature of this work they were cautious in interpreting the results.

The multiple regression analysis model of 15 linguistic and stylistic independent variables was statistically significant ($F(15, 72) = 11.802$, $P < .000$) in predicting readability for general audiences. However, only “Vocabulary” and “Main Point” statistically significantly contributed to whether annotators rated consumer health texts as readable for general audiences. The adjusted R^2 value indicated that 65% of variance within readability for general audiences was explained by the model. These results suggest that among the annotators and among all the linguistic and stylistic features, only two, “Vocabulary” and “Main Point," significantly predicted readability of a consumer health text.

For the second research question, the authors conducted a factor analysis of aggregated annotator responses to all 15 independent variables (n = 1320; 88 articles, each with 15 variables). A principal components analysis for all variables was tabulated, subject to a varimax rotation with Kaiser normalization. The number of factors was determined by an Eigenvalue > 1.0. Three factors accounted for 73% of the total variance, or 55%, 10% and 8% of the variance, respectively (Table 2). Due to the modest size of the data set, the authors were cautious in selecting a factor loading threshold and interpreting the results. Using a .50 threshold to determine a significant loading, factor 1 loadings included: lexical density, word repetition, words per package, cognitive load, discourse markers, bullets/lists, concept repetition, flow, and main point. Factor 2 loadings included word per sentence, sentence structure, and voice. Factor 3 loadings consisted of vocabulary and prior knowledge.

Table 2. Rotated Component Matrix(a) Component

Features	Factor 1	Factor 2	Factor 3
Vocabulary	.010	.243	**.870**
Lexical Density	**.693**	.444	-.043
Word Repetition	**.763**	.325	.141
Prior Knowledge	.153	.017	**.917**
Words per Sentence	.307	**.739**	.302
Words per Package	**.750**	.442	-.082
Sentence Structure	.247	**.775**	-.017
Cognitive Load	**.771**	.120	.363
Discourse Markers	**.891**	.155	.020
Personal Pronouns	.395	.490	.354
Voice	.041	**.702**	.206
Bullets/Lists	**.810**	-.006	.217
Concept Repetition	**.814**	.227	.221
Flow	**.883**	.288	.052
Main Point	**.769**	.179	.402

(a) Rotation converged in 5 iterations. Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.

Discussion

Only two features, "Vocabulary" and "Main Point," significantly predict whether the annotators rated consumer health texts as readable for general audiences. Traditional readability formulas incorporate syntactic (words per sentence) and semantic (vocabulary) features to predict readability. While the annotators verified familiarity with vocabulary as a predictor, they also found that effective communication of the main point is a significant attribute. These results may contribute to understanding consumer seeking and browsing health information online. For example, eye-tracking studies indicate that users typically scan a Web page for the "take-home" message and move on to another page if not found in a few seconds [9]. However, "ability to communicate the main point" is difficult to define operationally and measure.

All annotator assessments of the 15 features clustered within three statistically distinctive patterns or types of judgments used to assess the readability of consumer health text:

- Factor 1: discourse-level linguistic and stylistic features, reflecting an interest in document flow and general organization, both conceptually (lexical density) and structurally (bullets)
- Factor 2: "traditional" syntactic linguistic and stylistic features used in existing readability formulas [5] (ie, words per sentence) plus the effort in reading a passage (cognitive load)
- Factor 3: "traditional" semantic component of existing readability formulas (eg, vocabulary familiarity), where existing knowledge and term familiarity increase readability

Since readability is a multidimensional construct, all three perceptual orientations likely contribute to the assessment of consumer health texts. Hence, future work in developing methods to predict the readability of consumer health text should consider each cluster of features.

An important limitation of this study lies in its small annotator and document sample size. While the goal was to explore different ways of eliciting expert assessments of important features for each document, these results need to be verified with a larger data set.

The results of the multiple linear regression analysis in this study suggest that word- ("Vocabulary") and discourse-level ("Main Point") features contribute to general audience readability, unlike previous work which emphasize the role of syntactic features [3,8]. One practical implication is that these features might be given greater weight among the variables used to construct a more precise tool for predicting readability in the consumer health domain. The finding that the clusters include both linguistic and stylistic features indicates that text structure (sequencing/ordering of ideas or flow) with a logical transition between ideas and paragraphs or sentences (eg, discourse markers) is as important as vocabulary.

This study is the second in a series with the ultimate aim of building tools to predict the readability of consumer health text, thereby facilitating better matches to consumers with different reading abilities and information needs. The authors are currently evaluating how consumers with varying reading levels judge the expert assessments reported in this paper. Once the key features that predict readability are identified and validated, the research will focus on identifying quantitative measures of features or useful surrogates and integrating them into a predictive model of text readability in the consumer health domain.

Acknowledgements
This research was supported by the Intramural Research Program of the US National Library of Medicine and the US National Institutes of Health. The authors wish to thank Eunjung Kim, Harvard University, for her assistance with the statistical analyses.

Conflicts of Interest
None declared.

References
1. Pew Research Center. Trends 2005: Internet: the mainstreaming of online life. Washington, DC: Pew Research Center; 2005 Mar 20. [FREE full text]
2. Nielsen-Bohlman L, Panzer AM, Kindig DA, editors. Health Literacy: A Prescription to End Confusion. Washington, DC: The National Academies Press; 2004.
3. Chall JS, Dale E. Readability Revisited: The New Dale-Chall Readability Formula. Cambridge, MA: Brookline Books; 1995.
4. Gemoets D, Rosemblat G, Tse T, Logan R. Assessing readability of consumer health information: an exploratory study. Medinfo. 2004;11(Pt 2):869-73. Medline:15360936
5. Doak CC, Doak LG, Root JH. Teaching Patients With Low Literacy Skills. 2nd edition. Philadelphia: Lippincott Williams & Wilkins; 1996.
6. Graesser A, McNamara D, Louwerse M, Cai Z. Coh-Metrix: analysis of text on cohesion and language. Behavior Research Methods: Instruments, and Computers 2004 36:193-202.
7. MacDonald N, Frase L, Gingrich P, Keenan, S. The writer's workbench: computer aids for text analysis. IEEE Trans Comm 1982 COM-30;1:105-110.
8. Klare G. The Measurement of Readability. Ames, IA: Iowa State University Press; 1963.
9. Nielson J. F-shaped pattern for reading Web content. Alertbox 2006 Apr 17: http://www.useit.com/alertbox/reading_pattern.html [accessed 2006 May 22] Webcite: http://www.webcitation.org/5GLQTRx9p