Original Paper

# Assessing Consumer Health Vocabulary Familiarity: An Exploratory Study

Alla Keselman[1,2], PhD, MA; Tony Tse[1], PhD; Jon Crowell[3], MS; Allen Browne[1], PhD; Long Ngo[3], PhD; Qing Zeng[3], PhD

[1]National Institute of Health, National Library of Medicine, Bethesda, MD, USA

[2]Aquilent, Inc, Laurel, MD, USA

[3]Harvard Medical School, Decision Science Group, Brigham and Women's Hospital, Boston, MA, USA

**Corresponding Author:**
Alla Keselman, PhD, MA
Lister Hill National Center for Biomedical Communication, Room 7S713E
National Library of Medicine, NIH
8600 Rockville Pike, Bethesda, MD 20894, USA
Phone: +1 301 402 7851
Email: keselmana@mail.nih.gov

## Abstract

**Background:** Accurate assessment of the difficulty of consumer health texts is a prerequisite for improving readability. General-purpose readability formulas based primarily on word length are not well suited for the health domain, where short technical terms may be unfamiliar. We previously developed a regression model for predicting "average familiarity" with consumer health vocabulary (CHV) terms.Background:Accurate assessment of the difficulty of consumer health texts is a prerequisite for improving readability. General-purpose readability formulas based primarily on word length are not well suited for the health domain, where short technical terms may be unfamiliar. We previously developed a regression model for predicting "average familiarity" with consumer health vocabulary (CHV) terms.

**Objectives:** The primary goal was to evaluate the predictive ability of the CHV term familiarity model among actual consumers. Secondary goals were to explore the joint effect of demographic factors and familiarity and 2) surface level familiarity (ie, recognition) and an understanding of the underlying meaning (ie, conceptualization).Objective: The primary goal was to evaluate the predictive ability of the CHV term familiarity model among actual consumers. Secondary goals were to explore the joint effect of demographic factors and familiarity and 2) surface level familiarity (ie, recognition) and an understanding of the underlying meaning (ie, conceptualization).

**Methods:** Survey instruments for assessing surface level familiarity (45 items) and conceptual familiarity (15 items) were developed. All participants also completed a demographic survey and a standardized health literacy assessment, S-TOFHLA.Methods: Survey instruments for assessing surface level familiarity (45 items) and conceptual familiarity (15 items) were developed. All participants also completed a demographic survey and a standardized health literacy assessment, S-TOFHLA.

**Results:** Based on surveys completed by 52 consumers, linear regression suggests that predicted CHV term familiarity is statistically significantly correlated (Results: Based on surveys completed by 52 consumers, linear regression suggests that predicted CHV term familiarity is statistically significantly correlated (

*P* < .001) with participant scores on both the surface-level and conceptual familiarity surveys. Health literacy was a statistically significant predictor of surface-level familiarity scores (*P*< .001), and a marginally significant predictor of concept familiarity scores (*P* = 0.06). Educational level was not a significant predictor of either. Participant scores indicated that conceptualization lagged behind recognition, especially for terms predicted as "likely to be familiar" (*P*= .006).

**Conclusions:** This exploratory study suggests that the CHV term familiarity model is predictive of consumer recognition and understanding of terms in the health domain. Future validation studies will need to involve greater participant diversity and incorporate a more extensive selection of terms. Potential uses of such a model include readability formulas tailored to the consumer health domain and tools to "translate" professional medical documents into text that is more accessible to consumers.Conclusions: This exploratory study suggests that the CHV term familiarity model is predictive of consumer recognition and understanding of terms in the health domain. Future validation studies will need to involve greater participant diversity and incorporate a more extensive selection of terms. Potential uses of such a model include readability formulas tailored to the consumer health domain and tools to "translate" professional medical documents into text that is more accessible to consumers.

**KEYWORDS**

Consumer health vocabulary; patients; vocabulary; informatics; health education; readability; comprehension; health; evaluation studies

## Introduction

Improving the readability of online consumer health materials is an important area of e-health research. Studies suggest that health information on the Web is beyond the reading ability of average consumers [1,2]. Research on general literacy suggests that the readability decreases as the number of "difficult" words, those unfamiliar to the average reader, increases in a passage. Since familiarity correlates with education and literacy levels, "easy" terms are those that are familiar to many individuals who have lower reading skills. For example, the Dale-Chall readability formula incorporates a list of 3000 words and phrases (expressions) familiar to 80% of 4th-grade students in the United States [3]. However, because obtaining a comprehensive, empirically tested list of familiar words is difficult; many other existing readability formulas use average number of syllables per word as a surrogate for "word difficulty."

The notion of "general" word familiarity has been criticized: *"Counting words and syllables and consulting a grade-level word list are most likely not sufficient to determine how readable a text is"* [4]. Usage of word length is particularly ill-suited for the health domain, where short technical terms are likely to be, unfamiliar to consumers (eg, "apnea"). In addition, consumer health term familiarity is a complex concept that is more nuanced than existing binary measures, as summarized in [5]. Familiarity with health terms is viewed as a proxy for understanding the underlying concepts represented by terms. In cognitive science, a concept can be viewed as a set of slots that can be filled with characteristics describing a class of objects or events [6]. Clearly, the completeness and accuracy of conceptual knowledge exists on a continuum rather than a binary scale. Yet historically, health literacy studies do not distinguish between terms and concepts.

The authors had previously developed a regression model for predicting "familiarity likelihood scores" of consumer health vocabulary (CHV) terms using: (1) empirical data from user studies evaluating "Consumer-Friendly Display" names for medical concepts [7] and (2) term frequency counts from consumer health corpora [8]. The primary goal of this study was to develop and pilot test a simple methodology for validating the CHV familiarity predictive model by correlating it with two types of familiarity scores obtained from consumers: surface level familiarity (ie, recognition) and understanding of the underlying meaning (ie, conceptualization). The study also sought to describe the effect of demographic factors (including health literacy and education level) on the actual consumers' scores. Another goal was to relate characterize the relationship between the two types of familiarity scores, surface level familiarity and conceptualization. The following three hypotheses addressed the goals of the study:

1. Predicted familiarity likelihood level will have significant effect on consumers' surface-level term familiarity and their understanding of the underlying concept
2. Demographic factors, including but not limited to health and education level, will have significant effect on the both types of familiarity scores
3. Consumers' surface-level familiarity with terms will be greater than their understanding of the underlying concepts

## Methods

Participants: Consumers (n = 52) were recruited from Brigham and Women's Hospital; Demographic characteristics are shown in Table 1. According to Short Test of Functional Health Literacy in Adults (S-TOFHLA ) [9], 50 participants had adequate health literacy skills (scores in the 23-36 range out of 36), while two had marginal skills (scores from 17-22). Across all participants, scores ranged from 22 to 36, (mean = 33.04, SD=3.83).

**Table .** Demographic characteristics of the participants (n = 52)

| Demographic Variable | Distribution |
|---|---|
| Gender | 16 male, 36 female |
| English proficiency | 44 native speakers, 8 nonnative speakers |
| Highest education level | 2 below high school; 9 high school; 20 some college;<br>13 college; 8 graduate school |
| Age | 5 between ages 18-25; 13 between 25-39;<br>25 between 40-59; 9 over 60 |
| Race | 25 White, 13 Black, 8 Hispanic, 6 Other |
| Health literacy level<br>(STOFHLA scores) | 50 high health literacy (between 23-36);<br>2 moderate health literacy (between 17-22); |

Instruments: Surveys for assessing CHV term surface-level (45 items) and conceptual familiarity (15 items) were developed, piloted tested, and implemented. Their layouts were modeled on the Short Assessment of Health Literacy for Spanish-speaking Adults (SAHLSA) [10], which in turn is based on the Rapid Estimate of Adult Literacy in Medicine health literacy test for English speakers [11]. SAHLSA consists of 50 items, each with a "stem" or target term, "key" or semantically-related term, “distractor,” and a “don’t know” option to discourage guessing. Because SAHLSA is designed to measure both reading ability and comprehension, a correct answer requires the examinee to correctly pronounce the key and select the key term. However, since the goal was to measure familiarity with written health expressions and concepts explicitly using a self-administered tool (eg, via the Web), the SAHLSA requirement for examinees to pronounce each target expression was replaced with a second distractor (Figure 1).

**Figure .** Sample item from CHV familiarity tool

Renal: __heart __lung __kidney __don’t know
(stem) (2 distractors) (key)

Candidate CHV terms were selected from consumer health texts for three frequently visited MedlinePlus (http://medlineplus.gov) health topics: hypertension, back pain, and gastroesophageal reflux disorder (GERD). After all health-related expressions were manually extracted, they were submitted to a corpus-based expression familiarity predictive model. Each expression was categorized algorithmically into “predicted familiarity likelihood” levels as “likely” (0.8-1 familiarity score), “somewhat likely” (0.5-0.8 score), or “not likely” (0-0.5 score) to be familiar to consumers [8]. Five terms from each level were selected from each of the three health topic texts.

Two types of questions were developed:

- Surface-level familiarity: ability to associated written health terms with relevant, associated terms at a super-category level (eg, surgery   knife) (Figure 1).
- Concept Familiarity: ability to associate written terms with brief phrases describing the meaning or “gists” (eg, surgery   removing or repairing a body part).

The final test included surface-level familiarity items for all three health topics (questions 1 – 45), and concept familiarity items for GERD terms only (questions 46 – 60). The entire instrument is available in Multimedia Appendix 1.

**Administration and Scoring:** Participants first completed the demographics survey, followed by S-TOFHLA and CHV term familiarity survey, surface-level followed by concept familiarity. For scoring, each correct answer was awarded one point. Surface-level familiarity and concept familiarity scores were calculated separately.

## Results

**Descriptive Statistics: Mean Familiarity Scores:** Means and standard deviations of participants’ surface level term familiarity scores (both Total and GERD-only) and concept familiarity (GERD) are presented in Table 2. Maximum total surface-level familiarity score was 15; maximum GERD surface level familiarity and GERD concept score was 5.

**Table .** Mean surface-level and concept familiarity scores

| Predicted Familiarity Likelihood | Total Surface-Level Familiarity mean (SD) | GERD Surface-Level Familiarity mean (SD) | GERD Concept Familiarity mean (SD) |
|---|---|---|---|
| Likely | 13.80 (1.97) | 4.75 (0.81) | 3.83 (1.22) |
| Somewhat Likely | 12.92 (2.60) | 4.54 (1.02) | 3.94 (1.04) |
| Unlikely | 9.53 (3.44) | 3.42 (1.42) | 3.04 (1.31) |

**Testing Hypotheses 1 and 2:** Predictors of Surface-Level Term Familiarity: The independent variables Predicted Familiarity Likelihood Level, Gender, English proficiency, Highest Education Level, Age, Race, and Health Literacy Level (S-TOFHLA scores) were regressed onto the dependent variable, Surface-Level Term Familiarity Score. Linear regression found a statistically significant effect (P <. 001) of predicted familiarity likelihood level on surface-level term familiarity. Health literacy was a statistically significant predictor of surface-level familiarity ( P <. 001); English proficiency was marginally significant (P = 0.05); education level not at all (P = 0.15).

**Testing Hypotheses 1 and 2:** Predictors of GERD Concept Familiarity: GERD surface-level term familiarity and concept familiarity scores were computed. The range for both scores was 0-5, and corresponded to the number of GERD items on the survey. All independent variables (as described above) were regressed onto GERD concept familiarity score. Linear regression found statistically significant effects of predicted familiarity likelihood level (P = .009) and GERD surface-level score (P < .001) on concept familiarity scores. Health literacy level was marginally significant (P = 0.06).

**Testing Hypothesis 3:** Relating GERD Surface-Level and Concept Familiarity Scores: While previous regression analysis indicated that GERD surface-level familiarity score was a significant predictor of concept level familiarity, the concept familiarity consistently lagged behind surface-level familiarity at all three levels (see Table 2). Linear regression analysis of the effect of predicted familiarity likelihood level on the surface-level–concept familiarity gap was performed. For the overall model, the gap was statistically significantly different from 0 (P = 001). In addition, the gap is statistically significantly greater for terms predicted as “likely” then for terms “not likely” to be familiar (P = .006). The gap for terms predicted as “somewhat likely” is marginally greater than for those predicted “not likely” to be familiar (P = .07).

## Discussion

The findings of the study confirmed Hypotheses 1 and 3 and partially confirmed Hypothesis 2. Confirmation of Hypothesis 1 provided initial validity evidence for the CHV familiarity likelihood model [8] by demonstrating a relationship between predicted familiarity and two types of consumers' familiarity scores. The brief test used in this study should not be viewed as the final verdict on the feasibility of our statistical model. However, the survey instrument used in the study provides a methodological framework for additional validation studies. A fully validated CHV familiarity likelihood model may be used in multiple ways. For example, the algorithm could serve as a quick screening tool for determining "difficult" terms in consumer health texts and be incorporated into readability formulas tailored to the consumer health domain.

Partial confirmation of Hypotheses 2 and confirmation of Hypotheses 3 point to limitations of the model. Predicted familiarity likelihood scores are not adjusted for consumer groups with varying demographic characteristics, such as health literacy level, which affected the participants' term familiarity in this study. The predictive model framework also does not make a theoretical distinction between surface-level familiarity and conceptual understanding, and does not make provision for the possible uneven gap between the two. If this phenomenon of the uneven gap is validated in future studies, over-reliance on the predictive algorithm may result in the over-estimation of reader comprehension of consumer health texts.

While most of the study results corresponded to our research hypotheses, the lack of significant effects of most demographic variables, particularly educational level, is surprising and may be due to sampling bias. It is possible that uneven representation obscured any education effects 41 out of 52 participants had at least some college education. Note that education is a proxy for general literacy, which is only one component of health literacy [9]. Other components, such as healthcare experience and motivation, may have a much stronger effect on health term familiarity and needs to be explored in further research.

Follow-up work includes addressing the limitations of this study by extending the number of terms reviewed and obtaining a more balance sample of participants, representing a broad range of values on the demographic variables of interest (eg, healthcare experience). It is also essential to develop methods to explore consumer understanding of health concepts in-depth, as the current study only touches the surface of this important topic.

### Acknowledgments

This research was supported by the Intramural Research Program of the US National Library of Medicine, US National Institutes of Health (AK, TT, AB) and NIH grant R01 LM007222-05 (JC, LN, QZ). The authors thank Ilyse Rosenberg for her contribution to developing the instrument and Cara Hefner for assistance with the data collection.

### Conflicts of Interest

None declared.

### Multimedia Appendix 1

60-item questionnaire

[WinWord (.doc) file, XX KB - somedoc.doc ]

### References

1. Berland GK, Elliott MN, Morales LS, et al. Health information on the Internet: accessibility, quality, and readability in English and Spanish. JAMA 2001 May 23-30;285(20): 2612-2621. Medline: 11368735
2. Eysenbach G, Powell J, Kuss O, 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): 2691-2700. Medline: 12020305
3. Chall JS, Dale E. Readability Revisited: The New Dale-Chall Readability Formula. Cambridge, MA: Brookline Books; 1995.
4. McCray AT. Promoting health literacy. J Am Med Inform Assoc 2005 Mar-Apr;12(2): 152-63. Medline: 15561782
5. Chapman K, Abraham C, Jenkins V, Fallowfield L. Lay understanding of terms used in cancer consultations. Psychooncology 2003 Sep;12(6): 557-66. Medline: 12923796
6. Minsky M. A framework for representing knowledge. In: Winston PH, editor. The Psychology of Computer Vision. New York: McGraw-Hill; 1975: 211-277.
7. Zeng Q, Tse T, Crowell J, Divita G, Roth L, Browne AC. Identifying consumer-friendly display (CFD) names for health concepts. Proc AMIA Symp 2005: 859-63.
8. Zeng Q, Kim E, Crowell J, Tse T. A text corpora-based estimation of the familiarity of health terminology. Proc ISBMDA 2005: 184-92.

9. Baker DW, Williams MV, Parker RM, Gazmararian JA, Nurss J. Development of a brief test to measure functional health literacy. Patient Educ Couns 1999 Sep;38(1): 33-42. Medline: 14528569
10. Lee S-YD, Bender DE, Ruiz RE, Cho YI. Development of an easy-to-use Spanish health literacy test. Health Serv Res. In press.
11. Davis TC, Long SW, Jackson RH, et al. Rapid estimate of adult literacy in medicine: a shortened screening instrument. Fam Med 1993 25(6): 391-5. Medline: 8349060