Medical Informatics Europe 2002
G Surján et al. (eds)
IOS Press 2002

Readability of Patient Information on Diabetes on the Croatian Web Sites

Sanja KUSEC, Miroslav MASTILICA, Gordana PAVLEKOVIC, Luka KOVACIC

Andrija Štampar School of Public Health, Medical School, University of Zagreb, Rockefellerova 4, 10000 Zagreb, Croatia

Abstract. Many studies on readability have shown that patient information is frequently written at the reading level too high for the average population to understand. The aim of this study was to determine the reading level of patient information on diabetes found on the Croatian Web sites and written in Croatian language. Patient information on diabetes from 10 Croatian Web sites was tested for readability using the SMOG formula. The reading levels of the tested materials ranged from 11 to 15, which showed that these materials would not be understood by at least 80% of the Croatian adult population.

1. Introduction

With an estimated 300,000 Web users in Croatia, and a growing number of private, public and governmental Web sites related to health, patient education materials written in Croatian language and posted on the Web have become a popular mode of communication with the population at large.

Health literacy is defined as *'the degree to which individuals have the capacity to obtain, process, and understand basic health information and services needed to make appropriate health decisions'* [1]. However, for the population to act upon information and/or advice given on the Web, it is necessary that the written materials are adapted to the reading ability of that population.

Graber et al. showed that the reading level of Web users tends to be somewhat higher than that of patients in general [2], and the same could be assumed for the population in Croatia. However, the studies have shown that although the reading ability of patients varies widely, it is generally below the level of school completed [2], and most adults read a few grades (e.g. three to five grades in US adults) below their years of school completed [3].

Also, many studies have shown that most patient literature written in English fails to conform with the current standards of readability [4]. Therefore, it is hypothesized in this study that patient education texts written in Croatian language are also above the reading ability of the average Croatian population.

The aim of this study was to determine how difficult patient information found on the Croatian Web sites is to read in Croatian language.

2. Methods and Material

A standard methodology to test the readability of written materials is to apply one of the readability formulas. Nearly all 40+ readability formulas provide a reasonably accurate

estimate of grade level required to read, but not necessarily understand, a sample of text. Different formulas result in different scores; however, the results of various formulas correlate highly with one another [3].

The **S**imple **M**easure **O**f **G**obbledegook (SMOG) formula, introduced by McLaughlin in 1969, estimates the level of education required to understand the text and predicts 90-100% comprehension [3]. It has been used extensively to analyze health oriented literature. Proponents of this manual method of readability testing say SMOG is quick, consistent and easy to use.

However, since the readability formulas were created for English language in the first place, only later on were they modified in other languages (e.g. German, Spanish, French, Dutch, Chinese, Russian, Swedish, Vietnamese, etc.) [3]. For instance, a study by *Contreras et al.* showed that English is more readable than French, and French more readable than Spanish [5]. The SMOG scores have to be converted for languages other than English.

This study used the SMOG formula also for reasons of its easier application and adaptation to the Croatian language. The adaptation was based on the results of a comparison between the core vocabularies (100 top frequent words) of (American) English, based on Brown Corpus [6], and of Croatian language, based on the Croatian National Corpus [7]. The 10%-difference in the average word length and the 10%-polysyllabic (3+) structure observed in the Croatian core vocabulary required a correction in the definition of a polysyllabic word for Croatian language from 3+ syllable word to 4+ syllable word.

Many recent papers on health literacy present results of readability of patient information on such topics as diabetes mellitus, asthma and cancer [1). For the purposes of this study, the topic of diabetes was selected because considered of particular interest since the authors' institution conducted a field survey on nutrition in the 1970s and 1980s.

Patient information on diabetes on the Croatian Web sites was searched by entering the words *"šećerna bolest"*, a lay term used for diabetes in Croatian language, and using the CROSS search engine. CROSS (**Cro**atia **S**earch **S**ervice) is a search engine for searching the documents that reside on WWW servers in Croatia. It started in 1996, and presently is the official project of the Croatian Academic and Research Network (CARNet) [8].

There were 35 search results that were shown to an inexperienced Web searcher with secondary school education to select those that would be of interest to her while searching for information on diabetes. The selected 11 Web sites were then downloaded and printed for review. Review revealed that one Web site contained only Table of Contents of the material intended for physicians, and was excluded from the study. Patient information retrieved from the remaining 10 Web sites was then tested for readability using the SMOG formula.

The SMOG formula consists of 4 steps: 1) select 3 samples of 10 consecutive sentences each from different sections of text (at least 100 words total); 2) count the total number of 3+ syllable words in the 30 sentences; 3) take the square root of that number; 4) add 3. [3]. In case of a material with <30 sentences: 1) count all the polysyllabic words in the text, 2) count the number of sentences, 3) find the average number of polysyllabic words per sentence, 4) multiply that average by the number of sentences short of 30, 5) add that figure to the total number of polysyllabic words, 6) find the square root, and add the constant of 3 [9].

For the purposes of SMOG in this study, a sentence was defined as a string of words punctuated with a period. Bullets were considered one sentence only when an enumeration of word(s); when a complete sentence, each bullet was counted as one sentence. Hyphenated words were counted as one word. Numbers were written out to determine whether polysyllabic. In case of grammatical errors, words were counted as polysyllabic

upon correction. Apart from 4+ syllable words, all abbreviations and foreign words (Latin and English) were considered as polysyllabic words. Excluded from the testing were the introductory titles (because already obtained by search engine), the signatures, and the text in illustrations, tables, etc. if any. Calculations were done manually.

3. Results

The obtained readability levels of the patient information on diabetes from 10 Croatian Web sites tested by SMOG readability formula are listed in Table 1.

Table 1: Readability Levels of Patient Information on Diabetes from 10 Croatian Web Sites

SMOG grade*	No of sites (n=10)	Cumulative	% who would NOT understand
11	1 (10%)	10%	67.4
12	3 (30%)	40%	78.9
13	2 (20%)	60%	90.7
14	2 (20%)	80%	90.7
15	2 (20%)	100%	94.7
*SMOG grades 3-8 are equivalent to reading ability of people with primary level education, grades 9-12 are equivalent to those with secondary level education.			

The obtained SMOG grades ranged from 11 to 15, with the mean 13.1, and median 13.

The percentage of the Croatian population that would not understand the studied patient information on the Web was calculated according to the educational attainment of total adult population in Croatia.

The table shows that 60% of the materials would not be understood by at least 90% of the population. The percentages would be even higher if the reading ability was expressed at at least 3 years below the educational attainment.

4. Discussion

Reading level is expressed as the grade of education a reader needs to understand the information. In any given population, there will be a mix of reading ability [3]. For example, most Americans read at a sixth to eighth grade level, i.e. the reading level of the average American citizen is 7.5 [10]. The low literacy level of the British population requires the health literature to be written at a SMOG score ≤5 [11] so that it can be understood by most people. Even for sophisticated populations, materials designed to educate readers about health conditions and medical care require a low reading level. A high reading level (over grade 12) does not indicate the writing is appropriate for college educated readers but that the writing is complex and difficult to read [3].

In Croatia, according to the 1991 Census, the educational attainment of 54.7% of the total adult population (62.3% of the female adult population) is below or at the 8^{th} grade level [12]. Since the educational level does not necessarily translate into a corresponding level of reading or comprehension [1], at least a 3-year lower reading ability could be expected, as observed earlier for the American population [3].

The obtained SMOG readability grades for patient information on diabetes on Croatian Web sites ranged from 11 to 15, with the mean of 13.1. With reading ability of the population being below the years of school completed, we may estimate that at least 80% of the adult population in Croatia would not understand the information tested in this study. The 60% patient information studied (ranges 13-15) would be understood by less than 10% of the Croatian population. Taking into account that the Croatian language is less readable

than English, and that 4+ syllable words were taken as polysyllabic, these scores indicate to a very complex writing not adapted to the reading ability of the general population.

Furthermore, as limitations of the formula applied, we may state that strictly medical jargon, if a <4+ syllable word, i.e. easy to read, was not counted as polysyllabic although it is obviously not easy to understand. Since that would require a much detailed analysis that extends to the quality of information and the level of understanding, for the purposes of this study such an analysis was not done but was restricted to the level of readability only.

The high reading levels obtained in this study are consistent with the results obtained by *Payne et al.* [13] where SMOG scores range from 9 to 21, which indicates to very poor readability. Most leaflets in that study exceed the SMOG score of 10 or below, the recommended level for the information to be generally understood [13].

In the study by *Smith et al.* [11] SMOG scores of patient information leaflets range from 5 to 12, with the mean 8.66. The author considers SMOG readability grades 3-8 equivalent to reading ability of people with primary level of education, and grades 9-12 equivalent to those with secondary level education. The obtained scores, although much lower than in the study by *Payne et al.* are much higher than the recommended ≤5 for the low literate working British population.

Although the reading level of Web users tends to be somewhat higher than that of patients in general [2], and the same could be assumed for the Croatian Web users, with the growing number of Web users the adaptation of educational materials to a lower reading level might soon become a necessity.

Bearing in mind that readability formulas have a potential role in the evaluation of patient education materials which should be limited to serving as a guideline [10], the educational materials could be preliminarily tested for readability before being posted on the Web to avoid any complexity of writing.

Spiegel et al. wrote about the use of computers for the Flesch readability index but indicated also to the programming problems encountered in counting the words and syllables [14]. In Croatian language, morphology, syntax and semantics have all been observed as potential major problems in information retrieval and analysis [15]. So, further studies in this field are needed to enable implementation of human language technologies to the calculation of reading ease scores that would be done by word processing software for Croatian language.

5. Conclusion

The SMOG readability grades obtained in this study ranged from 11 to 15, and indicate to a very complex writing not adapted to the reading ability of the general population. Writers of patient information on the Web should be aware of the limited reading ability of their potential readers. So, policies might be needed to influence producers of such Web sites to adapt their material to the readers. These efforts may result in the creation of more readable materials. However, the only way to know for sure whether a material is suitable and effective is to test it with intended readers.

References

[1] Ratzan SC, Parker RM. Introduction - Health literacy (bibliography online). Bethesda, MD: National Library of Medicine, 2000. Available at: http://www.nlm.nih.gov/pubs/resources.html

[2] Graber MA, Roller CM, Kaeble B. Readability Levels of Patient Education Material on the World Wide Web. J Fam Pract 1999;48(1):58-61

[3] Smith S (1998). Readability Testing Health Information. Available at: http://www.prenataled.com

[4] Albert T. Written information for patients. Lancet 2000;356:434
[5] Contreras A, Garcia-Alonso R, Echenique M, Daye-Contreras F. The SOL formulas for converting SMOG readability scores between health education materials written in Spanish, English, and French. J Health Communication 1999;4(1):21-29
[6] Bujas Ž. O središnjoj jezgri američkoengleskog vokabulara. Jugoslavenska akademija znanosti i umjetnosti, Razred za filološke znanosti, Zagreb, 1986
[7] Moguš M, Bratanić M, Tadić M. Hrvatski čestotni rječnik. Zavod za lingvistiku Filozofskog fakulteta i Školska knjiga, Zagreb, 1999
[8] CROSS – Croatia Search Service. Available at: http://cross.carnet.hr/about_en.html
[9] http://oc.nci.nih.gov/services/HCPW/APPEN.htm
[10] Mathews PJ et al. The Use of Readability Formulas in Patient Education Materials. Paper presented at the American Association for Respiratory Therapy Summer Forum, Reno, NV, July 26-28, 1985
[11] Smith H, Gooding S, Brown R, Frew A. Evaluation of Readability and Accuracy of Information Leaflets in General Practice for Patients with Asthma. BMJ 1998;317:264-265
[12] Državni zavod za statistiku. Statistički ljetopis 2001. (Statistical Yearbook). Zagreb: Državni zavod za statistiku, 2001.
[13] Payne S, Large S, Jarrett N, Turner P. Written information given to patients and families by palliative care units: a national survey. Lancet 2000;355(9217):1792.
[14] Spiegel G, Campbell JJ. Measuring readability with a computer: What we can learn. Paper presented at the Meeting of the UCLA Conference on "Computers and Writing: New Directions in Teaching and Research", Los Angeles, CA, May 4-5, 1985
[15] Tadić M. Information Retrieval Meets Human Language Technology. Zagreb, 2000. Available at: http://www.carnet.hr/cuc/cuc2000/radovi/prezentacije/F/F4/F4_f.html