“Olá! Estás a ouvir-me?”
An exploration of health information seeking and likely comprehension of Internet-based health information in Portuguese and English
by Stephan Dahl, Lynne Eagle and Victor Souza

ABSTRACT: This paper focuses on health literacy and readability analysis as a guide for enhancing health information. Based on a selection of Internet-derived health and well-being materials, we show that the information available in both Portuguese and English is at a level likely to be too complex for the general population to understand. We also show a possible formula for adapting enabling the SMOG readability analysis tool so that it can be applied to information written in Portuguese and conclude the paper with a research agenda to help inform practice and policy debate.
Key words: Health Literacy; SMOG; Readability; Health Information

Olá! Estás a ouvir-me?
Estudo exploratório da informação de saúde e da sua compreensão, com base em fontes da Internet em português e inglês

RESUMO: Este artigo centra-se na literacia em saúde e na análise de legibilidade como um guia para a melhoria da informação de saúde. Com base numa seleção de materiais sobre a saúde e o bem-estar, provenientes da Internet, mostramos que a informação disponível, tanto em português como em inglês, está escrita num nível de complexidade provavelmente superior ao entendido pela população em geral. Ao mesmo tempo, propomos um ajuste ao modelo de análise de legibilidade SMOG, para que possa ser utilizado em informação escrita em português. Concluímos o artigo com uma agenda de pesquisa para informar a prática e o debate político.
Palavras-chave: Literacia em Saúde; SMOG; Legibilidade; Informação de Saúde
¡Hola! ¿Estás escuchándome?
Estudio exploratorio de la información de la salud y de su comprensión, con base en fuentes de Internet en portugués e inglés

RESUMEN: Este artículo se centra en la alfabetización en salud y en el análisis de la legibilidad como una guía para mejorar la información de la salud. Con base en una selección de materiales sobre la salud y el bienestar, procedentes de Internet, mostramos que la información disponible, tanto en portugués como en inglés, está escrita en un nivel de complejidad probablemente superior a la comprensión de la población en general. Al mismo tiempo, proponemos un ajuste al modelo de análisis de legibilidad SMOG, para que pueda ser utilizado en información escrita en portugués. Concluimos el artículo con una agenda de investigación para informar la práctica y el debate político.
Palabras clave: Alfabetización en Salud; SMOG; Legibilidad; Información de Salud

*Stephan Dahl*
sd@dahilat
PhD, University of Bedfordshire, UK. Adjunct Associate Professor; James Cook University, College of Business, Law & Governance, 1 James Cook Drive, Townsville, QLD 4811, New Zealand.
Doutorado, University of Bedfordshire, UK. Professor Associado Adjunto; James Cook University, College of Business, Law & Governance, 1 James Cook Drive, Townsville, QLD 4811, Townsville, Queensland, Australia.
Doctorado, University of Bedfordshire, UK. Professor Associado Adjunto; James Cook University, College of Business, Law & Governance, 1 James Cook Drive, Townsville, QLD 4811, Townsville, Queensland, Australia.

*Lynne Eagle*
lynn.eagle@jcu.edu.au
PhD, University of Auckland, New Zealand. Professor of Marketing; College of Business, James Cook Drive, Law & Governance, 1, Townsville, QLD 4811, New Zealand.
Doutorado, Universidade de Auckland, Nova Zelândia. Professora Catedrática de Marketing; James Cook University, College of Business, Law & Governance, 1 James Cook Drive, Townsville, QLD 4811, Nova Zelândia.
Doctorado, Universidade de Auckland, Nova Zelândia. Professora Catedrática de Marketing; James Cook University, College of Business, Law & Governance, 1 James Cook Drive, Townsville, QLD 4811, Nova Zelândia.

*Victor Gomes Lauriano Souza*
ju.souza@campus.fct.unl.pt
PhD candidate, Food Quality, Nova University of Lisbon, Portugal. Fellow, Federal Program Science Without Frontiers (Brazil), Nova University of Lisbon, Faculty of Science and Technology, Department of Science and Technology of Biomass, 2829-516 Caparica, Portugal.

Received in May 2017 and accepted in September 2017
Recibido em maio de 2017 e aceite em setembro de 2017
Recibido en mayo de 2017 y aceptado en septiembre de 2017
Estimates of the financial burden placed on societies and caused by preventable illnesses reaching £2,055 billion in the European Union alone (Kemp and Eagle, 2008), it is not surprising that social marketing and health promotion programs attempt to make even a small difference. There is thus considerable scope to improve population health through finding efficient and cost-effective means of conveying information as a foundation for subsequent behaviour change.

The critical role patients equipped with relevant, correct and well-understood information has been noted as one of the corner stones of modern patient-centred medicine (Mead and Bower, 2002) for well over a decade. It is therefore not surprising that providing patient information in the hope of empowering patients to actively part-take in their health is a frequent and important foundation of many health promotion and social marketing programs. Indeed, a great number of patients themselves tend to be interested in seeking out information about their health and well-being, with online health information playing a crucial role. Around 80% of U.S. Internet users have looked up health information online (Fox, 2011), with similar findings from Canada (Gibson, 2014). It is thus equitable to assume that online sources play a significant role in information seeking by patients in most developed countries. The information accessed online can include a wide variety of different types of sources: ranging from user-generated content (such as Wikipedia) to information provided by governments, health care providers and drug companies.

Usage and gratification theory can be used to explain information seeking online, and previous research has identified that online information seeking extends beyond merely information seeking. As Hou and Shim (2010) pointed out, many users seek to satisfy social and psychological needs in addition to seeking information online.

In an extensive review of the needs of online information seeking by primary care patients, Clarke et al. (2016) note the crucial role of seeking information online for patients in three areas: as an alternative to seeking health care advice, as pre-consultation information and empowerment strategy and post-consultation validation of advice received. They also note that the ease of access to information on the Internet is one of the most attractive features for patients.

While there is some debate about the usefulness specifically of pre-consultation information, doctors tend to view patients who accessed information online prior to consultation as a burden, despite pre-consultation information being linked to higher patient satisfaction (Massey, 2013). A possible explanation for this finding is potential confusion created by seeking information online. That is to say, a patient may be confused by, or misunderstand, the material accessed online, which results in an increased explanatory burden for the health care professional. As a result, there is an urgent need to understand the information seeking behaviour as well as the consequences of patients accessing health-related information online. This understanding can then be the foundation for developing more efficient strategies to aid health care profes-
sionals when responding to patient requests based on information sought online.

From a patient or online user perspective, effective seeking out and comprehension of online information requires substantial competencies: it requires the ability to conduct effective information searches, a degree of health literacy, and the need to evaluate the quality of information and the trustworthiness of information sources.

Health literacy itself is a complex set of skills: it has been defined by the U.S. Department of Health and Human Services Healthy People 2010 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” (Britt and Hatten, 2013, p. 2). While health literacy is linked to better health and treatment outcomes, the 2011 European Health Literacy survey found that around half of Europeans have limited health literacy (Sørensen et al., 2012). In Brazil, similar figures have been noted (Apolinario et al., 2014; Coelho et al., 2014). Specifically for online environments, e-health literacy is defined as the “ability to seek, find, understand, and appraise health information from electronic sources and apply the knowledge gained to addressing or solving a health problem” (Chen and Lee, 2014, p. 104).

E-health literacy is related to traditional literacy and numeracy, computer literacy, media literacy, science literacy and information literacy (Norman and Skinner, 2006), while equally related to existing health literacy in a conventional sense. In addition, health literacy has been noted to be context specific, with different knowledge and skills needed to prevent disease and maintain a healthy lifestyle versus the knowledge and skills needed to successfully navigate health services (Abel et al., 2014).

Beyond difficulties in understanding health-related information, people with low health literacy face far reaching consequences for themselves and providers: they are likely to use up more health care resources, resulting in higher overall health care costs. They are also more likely to attend emergency departments, have extended stays as inpatients and have inefficient mixes of health care services (Eichler, Wieser and Brügger, 2009).

A key skill required for successfully navigating, acquiring and understanding information online is digital literacy. Digital literacy is defined as “using digital technology, communications tools and networks to acquire and evaluate information, communicate with others and perform practical tasks” (OECD, 2012, p. 47). The link between (e-)health literacy and digital literacy levels is subject to some debate, although researchers have noted that high levels of digital literacy may be an enabler of health/e-health literacy, but low levels will be a barrier (Hu et al., 2012).

A comprehensive assessment of international literacy levels has been carried out by the OECD using a scale of 1-5 to report findings, where 5 is the highest level of literacy. For people to navigate work and life in an increasingly knowledge- and online-based environment, the OECD considers literacy levels of 3
or above as the minimum required (OECD, 2013). Level 3 is roughly equivalent to 9 to 10 years of formal secondary education. It should be noted, that a large portion of people in OECD countries do not reach the required level 3, although there are significant variations: 23.9% in Spain, 30.3% in Italy, 49.9% in the US, 49.5% in England, 41.7% in Australia, 51.4% in Germany (N.B. the report does not contain data from Portugal).

Readability

On the flip-side of literacy, readability is the ease with which a reader can understand a written text. Readability of a text depends on a variety of factors, including amongst others presentation of the text, complexity and variety of words used and sentence length. While text and layout cannot easily be classified, since the 1940s, researchers have developed different statistical formulae to evaluate the readability of materials and support writers in creating more readable texts.

A common feature amongst several of these formulae is the focus on polysyllable words: The Fry readability formula, the Flesch Reading Ease, the Flesch-Kincaid Grade Level, the Dale-Chall formula and the SMOG formula developed by McLaughlin all use polysyllable words as indicators of text complexity.

The most commonly used formula to assess readability in health communication is McLaughlin's SMOG formula (McLaughlin, 1969), short for Simple Measure of Gobbledygook, which estimates the years of formal education needed to be able to comprehend written material.

Since the development of the scale, it has been repeatedly validated and proven accurate; so much so, that the formula has been described as “the gold standard readability measure” (Fitzsimmons et al., 2010, p. 294).

The original formula required a sample of 30 sentences from a text and relied on the calculation of polysyllable words, defined as words that contain three or more syllables. Since its inception the formula has been refined and is now available as an online tool on the website of McLaughlin. The score (or grade) resulting from the formula indicates the years of formal education necessary to comprehend a scored text, e.g., a score of 9 would require 9 years of reading training, equivalent to the early years of high school.

It is important to remember, that reading, just as any other skills, requires ongoing practice to keep it at the highest level achieved in formal schooling. Consequently, the average adult reading skill is estimated to fall by 3-5 grades below the level expected at the completion of formal education if reading does not continue. That is to say, the reading skill of an adult who has completed high school, but who hasn’t continued to read with the same intensity is likely to fall to a grade 7 to 9 (Kemp and Eagle, 2008).

It is important to note, however, that different languages affect the readability scores or grades of all measures, including grades calculated by the SMOG formula. The reason for this variation can be found in the naturally occurring ave-
rage number of polysyllabic words and the average sentence length. For example, an English version of *The Little Prince* has 17707 words, of which 1023 are polysyllabic. A German translation of the book has 14646 words, of which 1293 are polysyllabic. The SMOG score of the English version of the book is 12.7, e.g., high school graduate level. The German version is 13.4, implying equivalent to early years of college education. In comparison, a Brazilian Portuguese version has 13155 words, 3238 of which are polysyllabic and 1591 sentences. The SMOG score of this version is 16.3, suggesting that comprehension of the text requires the equivalent to the reading level achieved when completing a university degree (all calculations by the authors based open access versions of *The Little Prince*). The question is thus if the formula can be successfully used in languages other than English.

Addressing the issue of different grades depending on the language, Contreras et al. (1999) evaluated the use of the formula for materials written in Spanish, English and French. They found that the SMOG formula is “systematically biased for text written in Spanish or French” (p. 27). Given the similarity of Portuguese and Spanish, it can therefore be assumed that the SMOG formula would be equally biased against materials written in Portuguese, as evidenced by the significantly higher grade of the Brazilian Portuguese version of *The Little Prince*. It is thus not advisable to compare rough SMOG grades between different languages.

However, Contreras et al. (1999) also demonstrated that, when taken as a guide, the SMOG formula was a valid means to characterise the readability of a text in different languages, provided that the final scores were adjusted according to a formula they called the SOL formula. The published SOL formula can be used for converting unadjusted SMOG grades obtained by examining material in French or Spanish to generally equivalent grades in English.

For texts written in Portuguese specifically, so far, no conversion formula has been published. Cavaco and Várzea (2010) used the SMOG formula to evaluate Portuguese-language information leaflets collected from pharmacies. The leaflets were first translated into English prior to being analysed. The average grade reported for the SMOG analysis was 11.13, significantly higher than the average reading skill of the population.

The purpose of this exploratory study is an evaluation of a range of online information and compare the readability of the available information between information available in English and Portuguese. This is an extension of an earlier study in which we evaluated the readability of materials published in English only. Rather than relying on translated material, which introduces the possibility of evaluating the readability of the translation rather than the original, we aim to work with the texts in their original language.

In order to achieve a comparable result between Portuguese and English-language grades, we therefore follow a two-step procedure. First we collect and analyse a selection of texts as they would be available to Internet users in an English-speaking country (Australia) and in a Portuguese speaking country.
(Portugal). We then use the SMOG formula to evaluate the readability of the selected texts. In a second step, we then calculated a conversion formula for comparable results between the texts written in Portuguese and English.

To evaluate existing health information available online, we used Google to locate five sources of health information linked to two health/well-being topics and two chronic conditions. These topics were: diet and exercise as health/well-being topics, and asthma and diabetes for chronic conditions.

To select appropriate sites and to allow for a comparison with an earlier study carried out on material available online in English (Dahl and Eagle, 2016), we selected the first five results a search using the site google.pt returned when searching from Portugal.

We followed the same procedure as the original study, which meant excluding any sites that were dictionary definitions (with the exception of Wikipedia), sales-based sites for herbal weight loss or asthma relief products, and sites that duplicated information such as Australian state and federal sites. We also excluded sites that were not in Portuguese to select the Portuguese samples.

Results

As can be seen from Tables 1 and 2, the average SMOG score differs vastly between sites written in English (12.8) and written in Portuguese (22.6). Previous research established that the SMOG formula can be used in languages other than English (Contreras et al., 1999), however, the scores or “grades” have to be adjusted according to the language. Contreras et al. (1999) suggest a formula to convert SMOG scores derived from Spanish and French sources into scores comparable to the original SMOG grades. For example, they suggest that for Spanish, the English grade can be derived by using the following formula: EN = -2.51 + 0.74*SP, where EN is the English grade and SP represents the SMOG score of the Spanish original text. Thus, if one assumes relative similarity of Spanish and Portuguese, the average score based on this formula would represent a grade of 14.21 in English.

In order to achieve a reliable estimate, we calculated an adjustment formula for Portuguese. The formula is based on the children’s book The Little Prince, which is readily available in both Portuguese (O Principezinho or O Pequeno Príncipe) and English. This follows the suggestion made by Contreras et al. (1999) to use works of literature as ways to find rudimentary conversion formulas, and their own analysis of The Little Prince.

Based on a sample of two chapters from the book, we calculated the relevant SMOG scores for both the English and the European Portuguese translation. Based on this, we calculated the approximate regression formula to convert scores from Portuguese (PT) to English (EN) grades. The resultant formula was EN = -6.48 + 0.94*PT. After adjusting the original average score for Portuguese-language texts using this formula, the average SMOG score of texts in Portuguese is 14.76. The estimated SMOG score reveals that texts written in Portuguese require a remarkably high level of reading skills. According to the
Table 1  SMOG analysis for general health and well-being Internet-based material

<table>
<thead>
<tr>
<th>Original Dahl and Eagle (2016) study</th>
<th>Portuguese Equivalent Sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organisation</td>
<td>URL</td>
</tr>
<tr>
<td>General health and well-being: a. Diet</td>
<td></td>
</tr>
<tr>
<td>Source</td>
<td>URL</td>
</tr>
<tr>
<td>----------------------</td>
<td>----------------------------------------------------------------------</td>
</tr>
</tbody>
</table>
Table 2 SMOG analysis for general health and well-being Internet-based material

<table>
<thead>
<tr>
<th>Organisation</th>
<th>URL</th>
<th>SMOG</th>
<th>SMOG</th>
<th>URL</th>
<th>Organisation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>General health and well-being: a. Diet</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>About.com</td>
<td><a href="http://index.about.com/%E3%81%A8%E3%81%AB%E3%81%8B%E3%81%8Fgclid=CNhje8g85MECFRUIvAodH74Aag&amp;am=broad&amp;q=information+about+diabetes&amp;an=google&amp;s&amp;askid=8a623177-5e3e-4891-9676-d5558de07725-0-ab_gsb&amp;dqi=&amp;qsrc=999&amp;ad=semD&amp;o=5946&amp;l=sem">http://index.about.com/とにかくgclid=CNhje8g85MECFRUIvAodH74Aag&amp;am=broad&amp;q=information+about+diabetes&amp;an=google&amp;s&amp;askid=8a623177-5e3e-4891-9676-d5558de07725-0-ab_gsb&amp;dqi=&amp;qsrc=999&amp;ad=semD&amp;o=5946&amp;l=sem</a></td>
<td>11.93</td>
<td>20.5</td>
<td><a href="http://holmesplace.pt/pt/dieta-para-quem-pratica-bodybuilding-a2918.html">http://holmesplace.pt/pt/dieta-para-quem-pratica-bodybuilding-a2918.html</a></td>
<td>Abbot FreeStyle Diabetes</td>
</tr>
</tbody>
</table>
### General health and well-being: b. Asthma

|-----------------------|-------------------------------------------------------------------------------------------------|-------|------|------------------------------------|-----------|
original formula, this would be approximately 14-15 years of formal education, equivalent to some college education. In comparison, texts written in English would require a less advanced reading skill (12 years of education), equivalent to a high school graduate.

Both scores give reasons for concern. The required reading level of texts in English of 12 is by far exceeding the likely reading level of the population. As previously pointed out, in the UK, the average reading level is likely to be approximately 9 (Eagle and Kemp, 2008).

No equivalent studies could be located for Portugal and Portuguese speaking countries. However, it is safe to assume that the required reading level of 14/15 to comprehend the material is likely to be substantially higher than significant parts of the general population. The average participation rate in Portuguese Higher Education is reported to be 35% (Governo de Portugal, 2010). That is to say, only a third of the population reaches the approximate level of education required to fully comprehend the texts analysed at the peak of their reading ability.

Discussion

The results indicate that information that is highly trusted, easily accessible and widely used by patients and users is written at a level of complexity that exceeds the reading ability of significant parts of the population. There is, consequently, a danger that large segments of patients using web-based information become confused or misunderstand materials they read.

This finding extends the findings of previous studies mostly based on traditional, paper-based information, which suggests that the nearly all of them are written at a level far exceeding the reading skills of the average population (Pires et al., 2015).

While some governments have made significant efforts to ensure accessibility and readability of health care information, it should be noted that many of the sites displayed are not government endorsed. Many of the top-displayed sites are private companies: such as pharmaceutical companies (MSD, Abbot, Bial), private health care providers (CUF, Lusíadas, in Portugal) or indeed media companies. Ensuring readability should therefore not only be a priority for governments, but these efforts should be extended to all sectors that provide health-related information.

It is important to acknowledge the role online information plays in health-related behaviour and decision making. This includes pre- and post-consultation information seeking, but also information being sought before any contact with health care professionals is being made.

While a growing body of research is focusing on medical information in the form of leaflets, package inserts and other written communication, it is important to expand the relatively weak theoretical foundations underpinning knowledge and of social and cultural factors that impact on whether, how and from what sources information is sought, interpreted and acted upon.

The results indicate that information that is highly trusted, easily accessible and widely used by patients and users is written at a level of complexity that exceeds the reading ability of significant parts of the population. There is, consequently, a danger that large segments of patients using web-based information become confused or misunderstand materials they read.
Future Research

The current research is exploratory in nature, and therefore has significant limitations, which, in turn, offer avenues of potentially fruitful future research. Regarding the availability of readability formulas specifically validated for languages other than English, we echo the call by Pies et al. for the need for further research. While the method we used is simple and offers face validity in this particular case, a more systematic development tool for the calculation of conversion formulae is desirable. This includes research not only in to the conversion of scores, but also further research into average literacy levels. Specifically, we note the high scores achieved in the OECD data for speakers of Roman languages other than Portuguese, all of which use naturally significantly more polysyllabic words in both spoken and written conversation.

In addition, future researchers should to focus on a wider range of medical conditions, both acute and chronic. Further research should also broaden the scope of potential information sources by including online discussion and support groups, specifically social networking sites. An important focus of future work must be to identify the combination of media channels used to gain information about drug treatments by patients, and compare the extent to which these channels influence patients’ beliefs, interactions with health professionals and health decisions.

Note

1. The Little Prince, first published in 1943 in the US, is a novella, the most famous work of French writer Antoine de Saint-Exupéry. Published in France in 1945, by Gallimard, as Le Petit Prince. (Editor’s Note)

References


MASSEY, P.M. (2013), Adoption and Use of Internet Technologies in Health Communication: Examining Disparities in Diffusion Patterns, Health Information Sources, and Patient-Provider Encounters. UCLA, Los Angeles.


