Contents lists available at ScienceDirect

American Journal of Infection Control

journal homepage: www.ajicjournal.org

Major Article

Readability of influenza information online: Implications for consumer health

Corey H. Basch EdD, MPH [a],[*], Joseph Fera PhD [b], Phillip Garcia BS [a]

[a] Department of Public Health, William Paterson University, Wayne, NJ
[b] Department of Mathematics, Lehman College, The City University of New York, Bronx, NY

Key Words:
Internet
Comprehension
Health literacy
Reading

Background: Influenza (flu) is pervasive and burdensome. The purpose of this study was to determine the readability levels of online articles related to flu.
Methods: Using the search term "influenza," the URL's of the first 100 English language Web sites were vetted for content to ascertain that the article met inclusion criteria. Five recommended readability tests were conducted using an online service to calculate readability. Overall, the analysis indicates that flu material found on the web is not being written at a level that is widely readable.
Results: None of the 100 sites included in the analysis received an acceptable score on all 5 assessments. One-sample independent t tests ($\alpha = 0.05$, df = 99) indicated that it is highly unlikely that flu Web sites are being written at the desirable level. Of the 100 sampled sites, 33 had a .com, 29 had a .org, and 22 had a .gov extension. Extension type did not play a role in readability level of these sites.
Conclusions: When creating content for the masses, health professionals should maximize their efforts by testing the readability as well as other factors that influence the likelihood that it will be understood.
© 2019 Association for Professionals in Infection Control and Epidemiology, Inc. Published by Elsevier Inc. All rights reserved.

Influenza (flu) virus causes contagious respiratory illness.[1] Complications can cause a myriad of consequences ranging from pneumonia to exacerbation of existing disease.[1] Although anyone can contract the flu, those at highest risk are those with existing medical conditions, pregnant women, children aged <5 years, and adults aged >65 years.[1]

Flu vaccine and proper hand hygiene can help prevent or minimize transmission.[1] Uptake of the flu vaccine remains suboptimal.[2] In the United States, vaccination coverage for adults during the 2017-2018 flu season was estimated to be 37.1%.[3] The annual number of flu cases in the United States is estimated to be between 9.2 million and 60.8 million since 2010.[4] Deaths related to flu have ranged between 12,000 and 56,000 since 2010.[4] Annually in the United States, there are an estimated 31.4 million outpatient visits and 200,000 hospitalizations related to the flu with an economic burden of approximately $87 billion.[5] Given the potential burden that accompanies the flu and the fact that all are at risk, finding ways to increase population-wide awareness about effective prevention and treatment methods is a high priority.

Nationally, over 70% of adult Internet users search online for health information in the United States.[6] In a study of 150 participant members of the lay public, lexical hints in written information were studied. Findings from this study indicated that the way in which articles are written influences perceptions. As such, public health professionals should be aware of the potential consequences of lexical hints in writing for the general public.[7] Another issue affecting the use and especially accessibility is the readability of written information.

Currently, 45 million Americans are unable to read above a fifth-grade reading level.[8] Although additional factors (such as access to information) can influence overall health literacy, understanding information is fundamental.[9] Public health professionals have recommended that materials be written at or below a sixth-grade reading level,[10] but we did not identify any published report of the readability of widely viewed Internet sites addressing flu. Therefore, the purpose of this study was to determine the readability levels of popular online articles related to flu.

METHODS

A sample of 100 Web sites hosting articles on flu were garnered using Google Chrome as a browser. Prior to conducting the search,

* Address correspondence to Corey H. Basch, EdD, MP, Department of Public Health, William Paterson University, 366 University Hall, Wayne, NJ 07470.
E-mail address: baschc@wpunj.edu (C.H. Basch).
Conflicts of interest: None to report.
Ethics approval and consent to participate: The human subjects committee at William Paterson University does not review studies that do not include human subjects.
Author contributions: C.H.B. conceived and designed the study. P.G. was the first coder. J.F. conducted the statistical analysis. C.H.B., P.G., J.F., wrote and edited the manuscript.

the browser was cleared. Using the search term "influenza," the URL's of the first 100 English language Web sites were vetted for content to ascertain that the Web site was relevant. If a Web site was merely a splash page, or did not have relevant content, it was excluded and the next was then assessed, and so on. Web sites that were splash pages, contained irrelevant information, were sponsored, or were merely an advertisement were excluded from the sample. A total of 17 Web sites were excluded. The basis for this exclusion is that these pages rarely eliminate the need to navigate to other web pages, being that relevant information pertaining to the focus of the Web site is not always presented directly on these platforms. According to Marshall,[11] a home page acts as the main area of the Web site, typically with the purpose to provide a brief overview of the Web site in its entirety. The home page invites the viewer to explore different offerings, menus, and tabs on the Web site as a way to draw the viewer in to stay. A home page will rarely provide detailed information pertaining to the focus of the actual Web site. Similarly, a splash page is another page the viewer is prompted with on first entrance to the web page. These pages are used to promote the service of the Web site, or to generalize the content available on the Web site. With that said, it is easy to use both terms interchangeably, as both are intended to provide simplified content and keep viewers fixated on the web page. Nevertheless, we can assume the content creators or Web site moderators have full discretion on whether useful information may or may not be provided.[11] Including splash/home pages with no relevant information to be a part of this sample has the potential to cause discrepancy in the readability scores overall and compromise the goal of the study.

The extension of each Web site (.gov, .com, .edu, etc.), was recorded for later analysis. MEDLINE recommends Readable.io[12] services for conducting readability tests, and therefore, was the chosen service for this project. This study used the following readability tests: Coleman-Liau Index (CLI), Flesch-Kincaid Reading Ease (FRE), Flesch-Kincaid Grade Level (FKGL), Gunning Fog Index (GFI), and the Simple Measure of Gobbledygook (SMOG).

It is important to note that these tests differ in some aspects such as the grading system used. FKGL and FRE both compute sentence and word length, but differ in scoring patterns.[13] FRE measures reading ease in that shorter sentences with easier words would score higher, higher scores infer that, in general, the text is more readable than those with a lower score.[13] Other tests, such as the FKGL and GFI, and CLI score by grade level similar to the US education system.[13] Of note, the GFI determines average words per sentence and the use of polysyllabic words, whereas CLI quantifies the number of letters in a word and number of words in a sentence.[13] Finally, the SMOG grade level uses a syllable counting system on a sample of text to establish the minimal education level needed to grasp the given material.[14]

Data analysis were performed using Microsoft Excel, version Office 2019 (Microsoft, Redmond, WA). For the purpose of analysis, parameters were set on what constituted an acceptable score and level of readability.[10] These were as follows: 80.0 or greater on the FRE and 6.9 or less on all other tests. This study did not undergo review by the institutional review board at William Paterson University, as it does not review studies that do not include human subjects such as this.

RESULTS

Overall, the analysis indicates that flu material found on the Web is not being written at a level that is widely readable. Table 1 shows 4 of the 5 readability tests classified a majority of Web sites as difficult, as opposed to easy or average, to read. Table 2 shows several statistics for the sampled Web sites. This table contains the mean readability scores for the Web sites on each of the assessments. The associated sample standard deviations are also given. To determine if

Table 1
Distribution of readability scores

Readability scores	Number of Web sites (n = 100)
FRE	
Easy (80-100)	0
Average (60-79)	19
Difficult (0-59)	81
FKGL	
Up to grade 7	14
Grades 7-10	44
Beyond grade 10	42
GFI	
Up to grade 6	13
Grades 6-10	40
Beyond grade 10	47
CLI	
up to grade 6	1
Grades 6-10	24
Beyond grade 10	75
SMOG	
Up to grade 6	0
Grades 6-10	19
Beyond grade 10	81

CLI, Coleman-Liau Index; FRE, Flesch-Kincaid Reading Ease; FKGL, Flesch-Kincaid Grade Level; GFI, Gunning Fog Index; SMOG, Simple Measure of Gobbledygook.

Table 2
Mean and standard deviation statistics

Readability test	Mean	Standard deviation	P value
FKGL	9.744	2.5091	7.46703E-20
GFI	11.72	13.896	0.000385193
CLI	11.61	2.5032	8.95306E-35
SMOG	11.67	2.0361	1.92891E-42
FRE	45.72	15.1866	4.02E-41

CLI, Coleman-Liau Index; FRE, Flesch-Kincaid Reading Ease; FKGL, Flesch-Kincaid Grade Level; GFI, Gunning Fog Index; SMOG, Simple Measure of Gobbledygook.

information on Web sites is being written at a recommended level, one-sample independent t tests ($\alpha = 0.05$; df = 99) were run. The P values obtained from these tests appear in Table 2. These P values, each considerably <.05, indicate that it is highly unlikely that flu articles posted on Web sites are being written at a readable level. The distribution of Web sites by extension type was noted. Of the 100 sampled sites, 33 had a .com, 29 had a .org, and 22 had a .gov extension.

Table 3 shows the average readability scores for the .com sites, .gov sites, and .org sites independently. None of the scores recorded meet the recommended readability level. Analysis of variance tests were performed to determine if extension type had a statistical difference on readability. The P values of these tests are also included in Table 4. Each P value is >.05, indicating that extension type does not play a role in readability level of these sites. Google often returns results 10 at a time. Therefore, sites 1-10 roughly correspond to page 1 results, sites 11-20 roughly correspond to page 2 results, and so on. Table 4 demonstrates the mean readability scores of the sites sampled when grouped by 10. The last column of this table shows the best average (in terms of readability) recorded in these groupings.

Table 3
Mean readability by extension

Readability test	.com	.org	.gov	P value
FKGL	10.41	9.66	9	.1151
GFI	15.39	10.38	8.46	.2042
CLI	11.93	11.7	11.5	.8229
SMOG	12.22	11.6	10.94	.0699
FRE	42.69	45.26	48.1	.4528

CLI, Coleman-Liau Index; FRE, Flesch-Kincaid Reading Ease; FKGL, Flesch-Kincaid Grade Level; GFI, Gunning Fog Index; SMOG, Simple Measure of Gobbledygook.

Table 4
Mean readability by search order

Search order	1-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90	91-100	Min/Max
FKGL	9.17	8.76	10.4	10.78	9.86	7.9	10.11	8.97	10.91	10.54	7.9
GFI	9.49	9.35	24.93	11.6	9.96	8.88	10.86	8.47	11.75	11.94	8.47
CLI	11.57	10.54	12.12	12.54	11.81	9.51	12.42	11.34	12.22	12.04	9.51
SMOG	11.20	11.72	12.34	12.5	11.49	10.15	12.12	10.83	12.58	12.64	10.15
FRE	47.77	49.84	41.97	39.82	44.21	54.94	44.13	47.61	41.64	45.29	54.94

CLI, Coleman-Liau Index; *FRE*, Flesch-Kincaid Reading Ease; *FKGL*, Flesch-Kincaid Grade Level; *GFI*, Gunning Fog Index; *SMOG*, Simple Measure of Gobbledygook.

Interestingly, it appears that page 5 results are the easiest to read based on these data.

DISCUSSION

The findings of this study suggest that online reading material on flu is, by and large, not being written at levels that are widely comprehendible. Given that the flu impacts so many individuals on an annual basis, a key component of educational efforts should be ensuring that individuals who are seeking information find it accessible. The intent of this study was not to compare the readability of flu information to other public health topics. The extent to which the results can be replicated for other topics remains to be seen, however, there are a variety of recent studies that suggest consistent findings that[15-16] material on the Internet is written at higher than desirable readability levels. Given the potential rapidity with which a highly pathogenic and virulent strain of flu may spread through the population, it is all the more necessary that widely accessed information is accessible to the majority of the population. Future studies can further assess the types of messages conveyed in online articles and whether those focused on a specific content area are more or less readable than others. For example, a readability comparison study conducted in Japan focused on information on the Internet that was in support or opposition of the flu vaccine. Findings of this study indicated that anti-flu vaccine content was written at a more understandable level than pro-flu vaccine content. Therefore, anti-vaccine sentiment may be more accessible to a widespread audience and have a negative impact on governmental vaccination efforts.[17]

This study is limited in that it solely focused on tests of readability. Additional tests could be conducted to give a broader picture of how the information provided might be received. For instance, tests of cohesion are not included in this report.[18,19] Also, these test do not measure how well the attention of the reader is held that requires different methodology.[20] It should also be noted that, although readability tools are recommended by trusted sources,[21] including advocating the use of online programs,[12] others mention their limitations.[20] Another factor to consider is the relatively small sample size and the fact that information on the Internet is in fluctuation and will likely not remain static over time. Therefore, the most widely identified articles are expected to change over time. This study does, however, fill a gap in the literature and could provide a platform for more comprehensive future research on this topic to demonstrate that the overwhelming majority of widely read sites are not written at a level that would be accessible to many people, both domestically and globally.

CONCLUSIONS

There is an abundance of information available to consumers. Many consumers are accessing information related to health to inform decisions for themselves, their family, and friends. Health agencies and researchers have noted the importance that material is written in a fashion that is as understandable as possible[21] as this influences health literacy,[22] which in turn can influence health outcomes.[23-26] Therefore, when creating content for the masses, health professionals should maximize their efforts by testing the readability as well as other factors that influence the likelihood that it will be understood.

References

1. Centers for Disease Control and Prevention. Key facts about influenza (flu); 2018. Available from: https://www.cdc.gov/flu/keyfacts.htm. Accessed April 2, 2019.
2. O'Halloran AC, Lu PJ, Williams WW, Bridges CB, Singleton JA. Influenza vaccination coverage among people with high-risk conditions in the US. Am J Prev Med 2016;50:e15-26.
3. Centers for Disease Control and Prevention. Estimates of influenza vaccination coverage among adults: United States, 2017−18 flu season; 2018. Available from: https://www.cdc.gov/flu/fluvaxview/coverage-1718estimates.htm#ref3. Accessed February 12, 2019.
4. Rolfes MA, Foppa IM, Garg S, Flannery B, Brammer L, Singleton JA, et al. Estimated influenza illnesses, medical visits, hospitalizations, and deaths averted by vaccination in the United States; 2016. Available from: https://www.cdc.gov/flu/about/disease/2015-16.htm. Accessed February 23, 2019.
5. Centers for Disease Control and Prevention Foundation. Flu prevention. Available from: https://www.cdcfoundation.org/businesspulse/flu-prevention-infographic. Accessed April 2, 2019.
6. Fox S. The social life of health information; 2014. Available from: http://www.pewresearch.org/fact-tank/2014/01/15/the-social-life-of-health-information/. Accessed April 2, 2019.
7. Mayweg-Paus E, Jucks R. Evident or doubtful? How lexical hints in written information influence laypersons' understanding of influenza. Psychol Health Med 2015;20:989-96.
8. Literacy Project Foundation. Illiteracy by the numbers. Available from: https://www.literacyprojectfoundation.org. Accessed February 23, 2019.
9. US Department of Health and Human Services. Office of Disease Prevention and Health Promotion. National action plan to improve health literacy. Washington (DC): Author; 2010.
10. McKenzie JF, Neiger BL, Thackeray R. Planning, implementing & evaluating health promotion programs: a primer. Upper Saddle River (NJ): Pearson; 2016.
11. Marshall T. Home page vs. splash page vs. landing page. Ardent creative; 2018. Available from: https://www.ardentcreative.com/home-page-vs-splash-page-vs-landing-page/. Accessed February 25, 2019.
12. National Institutes of Health. How to write easy-to-read materials; 2017. Available from: http://www.nlm.nih.gov/medlineplus/etr.html. Accessed February 12, 2019.
13. Readable.Io How readable is your writing? Available from: http://www.nlm.nih.gov/medlineplus/etr.html. Accessed February 12, 2019.
14. McLaughlin G. SMOG grading: a new readability formula. J Read 1969;12:639-46.
15. Basch CH, Ethan D, MacLean SA, Fera J, Garcia P, Basch CE. Readability of prostate cancer information online: a cross-sectional study. Am J Mens Health 2018;12:1665-9.
16. Basch CH, Fera J, Ethan D, Garcia P, Perin D, Basch CE. Readability of online information related to skin cancer. Public Health 2018;163:137-40.
17. Okuhara T, Ishikawa H, Okada M, Kato M, Kiuchi T. A readability comparison of anti- versus pro-influenza vaccination online messages in Japan. Prev Med Rep 2017;6:47-52.
18. McNamara DS, Graesser AC, McCarthy PM, Cai Z. Automated evaluation of text and discourse with Coh-Metrix. New York (NY): Cambridge University Press; 2014.
19. Crossley SA, Kyle K, McNamara DS. The tool for the automatic analysis of text cohesion (TAACO): automatic assessment of local, global, and text cohesion. Behav Res Meth 2016;48:1227-37.
20. Agency for Healthcare Research and Quality. Tip 6. Be cautious about using readability formulas; 2015. Available from: https://www.ahrq.gov/professionals/quality-patient-safety/talkingquality/resources/writing/tip6.html. Accessed April 2, 2019.
21. US Department of Health and Human Services. Public health service, National Institutes of health; making health communication programs work: a planner's guide. Available from: https://www.cancer.gov/publications/health-communication/pink-book.pdf. Accessed February 23, 2019.
22. Institute of Medicine (US) Committee on Health Literacy. Health literacy: a prescription to end confusion. editors. In: Nielsen-Bohlman L, Panzer AM, Kindig DA, eds. Health literacy: a prescription to end confusion. Washington (DC): National Academies Press (US); 2004. Available from: https://www.ncbi.nlm.nih.gov/books/NBK216029/. Accessed February 23, 2019.

23. Baker DW, Parker RM, Williams MV, Clark WS. The relationship of patient reading ability to self-reported health and use of health services. Am J Public Health 1997;87:1027-30.

24. Baker DW, Gazmararian JA, Williams MV, Scott T, Parker RM, Green D, et al. Functional health literacy and the risk of hospital admission among Medicare managed care enrollees. Am J Public Health 2002;92:1278-83.

25. Schillinger D, Grumbach K, Piette J, Wang F, Osmond D, Daher C, et al. Association of health literacy with diabetes outcomes. JAMA 2002;288:475-82.

26. Kalichman SC, Rompa D. Functional health literacy is associated with health status and health-related knowledge in people living with HIV-AIDS. J Acquir Immune Defic Syndr 2000;25:337-44.