Readability of Turkish Websites Containing COVID-19 Information

İbrahim Solak,[1] Betül Kozanhan,[2] Enes Ay[1]

[1]Department of Family Medicine, Konya Health Application and Research Center, University of Health Sciences, Konya, Turkey
[2]Department of Anesthesiology, Konya Health Application and Research Center, University of Health Sciences, Konya, Turkey

ABSTRACT

Objectives: This study aims to evaluate the readability levels of the information published on Turkish online sites about coronavirus-19 (COVID-19) as well as their contents.

Methods: The data of this descriptive study was obtained through http://www.google.com.tr, a frequently used search engine in Turkey. In May 2020, 464 websites accessed in the first 10 pages were evaluated using searches obtained using the keywords "Coronavirus," "COVID-19," and "What is the Coronavirus." The websites have been subsequently classified, analyzed, and divided into two groups according to the person, institution, organization, and so on uploading the information. The first group includes websites prepared by hospitals, government agencies, universities, and the second group includes websites prepared by news agencies and others.

Results: This study included 73 sites, 25 (34.2%) and 48 (65.8%) in the first and second groups, respectively. The median Ateşman and Bezirci-Yılmaz readability values of all sites included in the present study were 52.2 (46.2–57.3) and 11.9 (10.5–13.8), respectively. Moreover, the median number of syllables in the words, the median number of words in the sentence, and the median number of words in the sentence with 4 syllables or more were 2.8 (2.7–2.9), 12.4 (10.8–13.6), and 3.5 (3.0–4.2), respectively. Thus, the readability range of all sites included in the study is of medium difficulty according to Ateşman.

Conclusion: Consequently, the readability levels of Turkish websites about COVID-19 are determined to be of moderate difficulty and high school according to Ateşman and Bezirci-Yılmaz, respectively.

Keywords: COVID-19, readability, internet

INTRODUCTION

The coronavirus-19 (COVID-19) is a novel infectious disease in which the virus identified as severe acute respiratory syndrome coronavirus 2 (SARSCoV2) is the causative pathogen that has a close similarity to SARS coronavirus (SARS CoV) with SARS as the causative factor.[1] However, the pathogenesis of the virus is still unclear. Thus, the high rate of the spread of COVID-19 and its terminal impact leads to fear and anxiety in individuals. Therefore, access to accurate, reliable, and up-to-date information about COVID-19 during the pandemic is very important. Today, internet-based technologies are hugely acquiring worldwide attention. The worldwide web, the internet, is becoming increasingly more available for predicting, preventing, and monitoring new infectious diseases such as COVID-19.[2,3]

The term coronavirus has been reported to beat the top of the search lists worldwide with the

COVID-19 outbreak reaching global dimensions, according to Google trends analysis data.[4] In Turkey, like in the whole world, the most frequently asked questions in Google-search engine were "What is the coronavirus?," "How many people died from coronavirus today?," "How long does it take to get the coronavirus test results?," "Where did the coronavirus outbreak?," "What are the methods of protection and vaccines against coronavirus?," "What are the suggestions for reducing virus infection?," and "What are the symptoms of coronavirus?".[5]

The internet has great potential to spread information related to health and well-being.[6] Web-based information affects how patients comply with the recommendations, clinical diagnoses, and treatment regimens recommended by healthcare professionals.[7,8] However, as a source of health information, the internet is unregulated and varies in quality, reliability, and reader accessibility. Many web pages contain false or dubious information that can yield more harm than benefit. Misinformation that is spread over the internet can prevent the communication of health institutions and professionals with the general population and decrease compliance with the recommendations of the National Pandemic Coordination Board such as social distancing.[9]

Whereas the pressure and urgency to publish information in a short period during times of global crises such as during a pandemic is evident, understanding relevant information and literature on health issues is of the most importance. Moreover, the basic requirement to effectively understand and use health information is through the accessibility and readability of the information presented. Readability is a concept associated with the reading difficulty of a text objectively measured with mathematical formulas specific to a certain language.[10] Approximately 40.0% of the American population is insufficiently health-related literate. In Turkey, the situation is similar as revealed in a 2014 study which determined that 64.6% of the Turkish population were insufficiently health-related literate.[11] Therefore, messages regarding risks in health literature and information should contain accurate/scientific information and should be easily understood by the majority of the population.[12,13] Health professionals recommend that these informational texts should be at the reading level of the 6th grade to yield the most benefit to the general public.[14]

Many studies have evaluated the readability of information available on the internet, depending on different health conditions. Thus, the present study evaluates the readability levels of the information published on Turkish online sites about COVID-19 as well as its contents.

METHOD

The data of this descriptive study was obtained through http://www.google.com.tr, a frequently used search engine in Turkey. In May 2020, 464 websites accessed in the first 10 pages were evaluated using searches obtained from the keywords "Coronavirus," "COVID-19" and "What is the Coronavirus."

Sites containing information of fewer than 10 sentences, chat forums, other forums, commercial blog sites, and sites containing only pictures, tables, and videos were excluded. The informational texts on these sites were copied to Microsoft Word, and educational titles, author information, site URL, addresses, and links were deleted to prevent the negative impact on readability results. The websites have been subsequently classified, analyzed, and divided into two groups according to the person, institution, organization, and so on uploading the information. The first group includes websites prepared by hospitals, government agencies, universities, and the second group includes websites prepared by news agencies and others.

Readability Measurement: In calculating the readability level of the informative texts on the websites, the formulas of Ateşman and Bezirci-Yılmaz were used to determine the readability level of Turkish text. First, the internet-derived texts were edited using Microsoft Word. Then, the average number of words (average sentence length), syllables (average word length), and words of four syllables and more were calculated manually using the Excel program.[10] Lastly, the data obtained was thus transferred to the software to calculate the readability values determined using the formulas of Ateşman and Bezirci-Yılmaz.[15,16]

Ateşman Readability Formula: The Ateşman readability formula was developed by adapting the Flesch Reading Ease Formula to Turkish based on word and sentence length.[16] In the Ateşman formula, the readability level of the text is determined as very easy, easy, medium, hard, and very difficult the scores are between 90 and 100, 70 and 89, 50 and 69, 30 and 49, and 1 and 29, respectively.

The Ateşman readability formula: Readability score = 198.825−40.175 × (total syllable/total word)−2.610 × (total word/total sentence)

Bezirci-Yılmaz Readability Formula: The Bezirci-Yılmaz readability formula was developed in 2010 based on sentence lengths in the texts and the number of syllables in the words considering the features of the different formulas developed and the statistical properties of the Turkish language.[15] The numbers of the syllables in the words are

calculated by multiplying them with their specific numbers, and the level of readability is formulated as:

NRW= $\sqrt{ANW}$*((H3*0.84)+(H4*1.5)+(H5*3.5)+(H6*26.25))

where NRV is the new readability value, ANW is the average number of words, H3 is the average number of three-syllable words, H4 is the average number of four-syllable words, H5 is the average number of five-syllable words, and H6 is the average number of words with six or more syllables.

As the lengths of the sentences in the texts increase, the readability difficulty of the texts increases as well according to this formula. Likewise, the increase in the number of syllables in the words also makes it difficult to read the word and thus often that of the sentences. The results obtained from this formula explained to which grade levels a text addresses according to the Turkish educational system.[15]

In the Turkish educational system, primary, secondary (high school), tertiary, and higher academic education levels are divided into grades 1–8, 9–12, 12–16, and 16, respectively.

Content Evaluation of Texts: The COVID-19 guide prepared by the Turkish Ministry of Health of the Republic of Turkey for public information was taken as the basis to assess the elements of informative content in the internet-based text accessed via specific keywords.[17] The texts were examined according to the answers that these internet-driven texts provide to any of the questions in the guide.

The Statistical Package for the Social Sciences, version 21.0 (IBM Inc., Armonk, NY, USA) was used for data analysis. The categorical and numerical data used in this study were shown by using frequency and percentage values as well as the median and minimum–maximum values, respectively. Whether the study groups were distributed normally or not was tested by using the Shapiro–Wilks and Kolmogorov–Smirnov tests. Moreover, the Mann–Whitney U-test and the Chi-square test were used in this study to compare the numerical data between the independent groups and the categorical data, respectively. In all analysis results, p<0.05 was considered significant.

RESULTS

This study included 73 sites divided into 25 (34.2%) and 48 (65.8%) in the first and second groups, respectively. The median Ateşman readability value of all sites included in the present study was 52.2 (46.2–57.3) and the median Bezirci-Yılmaz readability values was 11.9 (10.5–13.8) years. Moreover, the median number of syllables of the words, the median number of words in the sentence, and the median number of words in the sentence with 4 syllables or more were 2.8 (2.7–2.9), 12.4 (10.8–13.6) and 3.5 (3.0–4.2), respectively.

A statistically significant difference exists between the Ateşman readability values, Bezirci-Yılmaz readability values, number of words in the sentences, and number of words in the sentence with four syllables and above (p=0.006, p=0.011, p=0.016 and p=0.034, respectively). However, no statistically significant difference could be determined between the mean number of syllables of the words (p=0.098). Comparison of the readability level of the groups are summarized in Table 1.

The readability range of all sites included in the study was of medium difficulty according to Ateşman. The websites included in the study were divided into two groups according to the Ateşman readability intervals. Moreover, a statistically significant difference exists between the readability intervals of the study groups according to Ateşman (p=0.003). Evaluation of the groups readability intervals according to Ateşman are summarized in Table 2.

When the contents of the sites included in the study were examined, 66 (90.4%), 67 (91.8%), 66 (90.4%), 27 (37.0%), 33 (45.2%), 59 (80.8%), and 45 (61.6%) contain information about the definition, symptoms, forms of transmission,

Table 1. Comparison of the readability level of the groups

	Group 1 (n=25)	Group 2 (n=48)	p
Ateşman readability value	47.7 (45.3-54.4)	53.0 (48.6-58.2)	0.006
Bezirci-Yılmaz readability value	13.3 (11.4-15.0)	11.5 (10.1-13.2)	0.011
Mean number of syllables	2.8 (2.7-2.9)	2.8 (2.7-2.8)	0.098
Number of words in the sentences	13.2 (11.9-14.9)	11.5 (10.5-13.3)	0.016
Words consisting of four or more syllables	4.0 (3.1-4.3)	3.4 (2.9-3.9)	0.034

All data are expressed as median (min-max).
Mann-Whitney U Test.

Table 2. Evaluation of the groups readability intervals according to Ateşman

Readability level	Group 1	Group 2	Total	p
Very difficult and difficult	15 (55.6)	12 (44.4)	27 (100.0)	0.003
Medium difficulty, easy and very easy	10 (21.7)	36 (78.3)	46 (100.0)	

All data were presented as n (%).
Chi-Square Test.

groups at risk, diagnostic procedures, methods of prevention, and treatment of COVID-19, respectively. When site contents of the groups included in the study were compared, no statistically significant differences could be determined in terms of COVID-19 definition, symptoms, transmission mode, risk groups, diagnosis, protection, and treatment between the two groups (p=0.410, p=0.666, p=0.087, p=0.221, p=0.127, p=0.354, and p=0.458, respectively).

DISCUSSION

On March 11, 2020, the World Health Organization (WHO) declared COVID-19 as a globally spread epidemic.[18] Moreover, on June 3, 2020, WHO reported that 6.242.974 patients have been diagnosed with COVID-19 worldwide resulting in 378.485 mortalities.[19] Currently, the number of people affected is still rapidly increasing worldwide.

In all infectious diseases, transmission prevention is the way to not become infected. The most important element of preventing infection is to inform society about the epidemic and the factors leading to infection. In the last two decades, sources to acquire information in the field of health have rapidly changed. Whereas people used to consult doctors and healthcare providers for health-related information, the internet has become the most common source for people while taking a medical decision today.[20] This situation changes in favor of the internet with its widespread use. Therefore, the readability and legibility of health-related information prepared on the web sites are very important.

The Human Development Report, published in 2019, reported an average of 7.7 years of education in Turkey.[21] In the present study, the mean readability values of the web-based informative texts examined were medium difficulty and high school according to Ateşman and Bezirci-Yilmaz, respectively. In other words, only those who have been educated for 11.9 years can comprehend these texts in total. This value is much higher than the average educational level of the Turkish population. However, this score is far above the educational level of the common people. This situation re-

veals that the information on the websites about COVID-19 is not comprehended by the Turkish population.

Worrall et al. reported that only 17.2% of the websites prepared about COVID-19 in the English language are at a universally readable level.[22] Moreover, Basch et al. reported that the readability of English websites about coronavirus exceeded the 10th-grade reading level. This level was above the reading level of an average American.[23] The results of the present study are in line with the research conducted about the readability of information on the websites about COVID-19 in English-language literature.

The present study is the first to examine Turkish websites about COVID-19. Moreover, some studies exist examining the readability of websites that contain medical topics in the Turkish language. Solak examined websites containing information about colorectal cancer and reported the median readability level value according to Ateşman and Bezirci Yılmaz as 50.81 and 12.32, respectively.[24] In addition, Kozanhan et al. evaluated the readability of patient information texts presented on the websites about anesthesiology and determined the readability of these to be within the range of difficult and undergraduate level according to Ateşman and Bezirci Yılmaz, respectively, according to the Turkish educational level.[10] The readability values of the texts on the websites that contain information about autism spectrum disorder conducted by Kara et al. were 43.72 and 14.68 according to Ateşman and Bezirci-Yılmaz, respectively.[25] Likewise, in their study on specific learning difficulties, Kara et al. reported the readability values of 46.01 and 13.82 according to Ateşman and Bezirci-Yılmaz, respectively.[26]

The readability levels of the texts prepared on the internet in the relevant literature are quite above the education and academic level of the general public of this study. The average sentence length in the Turkish language was calculated as nine to 10 and 10 to 11 words according to the formulas of Ateşman and Bezirci-Yılmaz, respectively. In both formulas, the average word length was calculated as 2.6 syllables.[15,16] In the present study, the average number of words in texts and the average number of syllables in the words are higher than these values. Thus, its readability is harder than usual.

The sites in the first group (state institutions, universities, and hospitals) are harder to read compared to the second group when the readability values of the groups included in the study were analyzed. The harder readability of the sites prepared by these institutions, where people can reach reliable information, makes it difficult for people to

access correct information. Besides the readability of the text provided on the internet, their sufficiency and correctness an equally important. Thus, sufficient information about COVID-19 is contained in the websites included in the study.

CONCLUSION

As a result of the present study, the readability levels of websites prepared in Turkish about COVID-19 are determined to be of moderate difficulty and high school according to Ateşman and Bezirci-Yilmaz, respectively. In other words, 11.9 years of education are necessary for these texts to be understood considering the Turkish educational system. The texts are to be written with sentences consisting of few words as possible and words that contain as few as possible syllables for the information on the websites to be better understood.

Disclosures

Peer-review: Externally peer-reviewed.

Conflict of Interest: None declared.

Ethics Committee Approval: Approval for the study was granted by the Hamidiye Scientific Research Ethics Committee, University of Health Sciences (Approval date: Sept 21, 2020 and Approval number: 2020/314).

Authorship Contributions: Concept – I.S., B.K.; Design – I.S., B.K.; Supervision – I.S., B.K.; Materials – I.S., E. A.; Data collection &/or processing – I.S., E. A.; Analysis and/or interpretation – I.S., E. A.; Literature search – I.S., E. A.; Writing – I. S., B. K.; Critical review – I. S., B. K., E.A.

REFERENCES

1. Yang X, Yu Y, Xu J, Shu H, Xia J, Liu H, et al. Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. The Lancet Respir Med 2020;8(5):475–81.

2. Effenberger M, Kronbichler A, Shin JI, Mayer G, Tilg H,Perco P. Association of the COVID-19 pandemic with internet search volumes: a Google TrendsTM analysis. IJID 2020;95:192–97.

3. Yang T, Gentile M, Shen C-F, Cheng C-M. Combining point-of-care diagnostics and internet of medical things (IoMT) to combat the COVID-19 pandemic. Diagnostics (Basel) 2020;10(4):224. [CrossRef]

4. Anadolu Agency. Coronavirus dominates Google trends during lockdown. Available at: https://www.aa.com.tr/en/latest-on-coronavirus-outbreak/coronavirus-dominates-google-trends-during-lockdown/1832037. Accessed May 3, 2020.

5. Google Trends. https://trends.google.com/trends/. Accessed May 2, 2020.

6. Berland GK, Elliott MN, Morales LS, Algazy JI, Kravitz RL, Broder MS, et al. Health information on the Internet: accessibility, quality, and readability in English and Spanish. JAMA 2001;285(20):2612–21. [CrossRef]

7. Lu X, Zhang R, Wu W, Shang X, Liu M. Relationship between internet health information and patient compliance based on trust: empirical study. J Med Internet Res 2018;20(8):e253.

8. Lu X, Zhang R. Impact of physician-patient communication in online health communities on patient compliance: cross-sectional questionnaire study. J Med Internet Res 2019;21(5):e12891. [CrossRef]

9. Farooq A, Laato S, Islam AKM. Impact of online information on self-isolation intention during the Covid-19 pandemic: cross-sectional study. J Med Internet Res 2020;22(5):e19128. [CrossRef]

10. Sena T, BASIM P. A new perspective on readability and content assessment of patient information texts published on the internet sites on lymphedema. Journal of Current Researches on Health Sector 2018;8(2):303–14.

11. Edworthy J, Hellier E, Newbold L, Titchener K. Passing crisis and emergency risk communications: The effects of communication channel, information type, and repetition. Appl Ergon 2015;48:252–62. [CrossRef]

12. Basch CH, Basch CE, Redlener I. Coverage of the Ebola virus disease epidemic in three widely circulated United States newspapers: implications for preparedness and prevention. Health Promot Perspect 2014;4(2):247.

13. McKenzie JF, Neiger BL, Thackeray R. Planning, implementing, and evaluating health promotion programs: a primer. 6th ed. San Francisco: Benjamin Cummings; 2012.

14. Bezirci B, Yılmaz AE. A software library for measurement of readability of texts and a new readability metric for Turkish. DEÜ FMD 2010;12(3):49–62.

15. Ateşman E. Measuring readability in Turkish. AU Tömer Language Journal 1997;58(2):171–74.

16. T. C. Ministry of Health New Coronavirus Disease Booklet. Available at: https://covid19bilgi.saglik.gov.tr/depo/Kitapcik/COVID-19_YENI_KORONAVIRUS_HASTALIGI_KITAPCIK_A6.pdf. Accessed May 3, 2020.

17. Ranney ML, Griffeth V, Jha AK. Critical supply shortages—the need for ventilators and personal protective equipment during the Covid-19 pandemic. N Engl J Med 2020;382(18):e41.

18. WHO. Coronavirus Disease (COVID-19) Dashboard. Available at: https://covid19.who.int/. Accessed May 3, 2020.

19. Koch-Weser S, Bradshaw YS, Gualtieri L, Gallagher SS. The Internet as a health information source: findings from the 2007 Health Information National Trends Survey and implications for health communication. J Health Commun 2010;15(sup3):279–93. [CrossRef]

20. United Nations Development Programme. Human Development Reports http://www.hdr.undp.org/en. Accessed May 6, 2020.

21. Worrall AP, Connolly MJ, O'Neill A, O'Doherty M, Thomton KP,

McNally C, et al. Online Readability of COVID-19 Health Information. BMC Public Health 2020;20:1635. [CrossRef]

22. Basch CH, Mohlman J, Hillyer GC, Garcia P. Public health communication in time of crisis: readability of on-line covid-19 information. Disaster Med Public Health Prep 2020;14(5):635–7. [CrossRef]

23. Solak M. Kolorektal kanser hakkında bilgi içeren internet sitelerinin okunabilirliği. Journal of Harran University Medical Faculty 2019;16(3):509–13. [CrossRef]

24. Kozanhan B, Tutar MS. Anesteziyoloji alanında internet sitelerinde sunulan hasta bilgilendirme metinlerinin okunabilirliklerinin değerlendirilmesi. Turkiye Klinikleri J Anest Reanim 2017;15(2):63–70. [CrossRef]

25. Kara A. Evaluation of the readability of online texts related with autism spectrum disorder. Exp Biomed Res 2019;2(4):136–42.

26. Kara A, Polat H. Assessment of the readability of online texts related to specific learning disorder. Med Science 2020;9(1):114–7. [CrossRef]