Readability of Websites for the Novel Coronavirus Disease (COVID-19): A Case Study of India

Diksha Dani and Gaurav Agrawal

Abstract The World Wide Web has become one of the most important resources for the large part of the population in searching healthcare information. The available COVID-19 information on Websites should be reliable and accurate in order to spread awareness. This has become all the more important in times like COVID-19, pandemic situation when it is necessary to create awareness about the disease to large population in a country like India and also provide updated information and status of the situation. However, readability poses barrier for many people. In this paper, readability of online information related to the novel coronavirus disease (COVID-19) is evaluated using online readability tool. It was found that the readability of content addressing COVID-19 is satisfactory. But the absence of multilanguage option in majority of Websites under study forms another barrier in information dissemination.

Keywords COVID-19 · Readability · WebFX tool

1   Introduction

India is a country with the second largest population in the world. The COVID-19 disease that has brought the world to its knees is highly contagious, and social distancing has been suggested as one of the measures to prevent its spread. Owing to India's large population, it is difficult to maintain social distancing; consequently, the disease has a higher probability of spreading faster. Also, due to social media platforms, false information and rumors may spread fast. So, communicating the correct updated information about preventive measures to the maximum possible people is a massive task. The Internet has proved to be quite a boon in such situations. But, there is another issue with this task. India is a land of diverse languages and dialects, and the majority of the population lives in rural areas. Therefore, it is equally important that the information available online on highly communicable disease like COVID-19 is to be conveyed in such a manner that the majority of the population

D. Dani (✉) · G. Agrawal
Inderprastha Engineering College, Ghaziabad, India

understands. Since, due to the spread of digital information deep even in remote areas, the Internet has become the primary source of information regarding awareness and treatment of patients, it is essential to evaluate the readability levels of the information presented by the online Websites. This study evaluates the readability of information available online related to the COVID-19 pandemic in Indian Websites.

2 Related Work

There are many studies on the readability of healthcare material available online. Grose et al. [1] evaluated the readability and quality of online information on septoplasty. The authors concluded that most online education materials on septoplasty have significant quality and reliability issues and severely high reading levels. Likewise, in another study [2], online education materials on congestive heart failure for patients were evaluated for readability. Out of 70 Websites, only 5 met the recommended sixth-grade readability level.

However, there are not many papers explicitly about the COVID-19 pandemic on the Internet. The authors of the paper [3] assessed the readability of information available over the Internet about the novel coronavirus disease (COVID-19) to determine compliance with reading level standards. They found that online content about COVID-19 is complicated to read and understand, as advocated by the American Medical Association (AMA) and the United States Department of Health and Human Services. Kruse et al. conducted another study linked to COVID-19 reading [4]. The researchers assessed the content of the Website, readability of the text present on site, and quality of Web-based PEMs from U.S. academic medical centers on COVID-19. They discovered that the mean readability was greater than the suggested sixth-grade reading level set by the National Institute of Health and the U.S. Department of Health and Human Services. Worral et al. [5] evaluated the readability of COVID-19-related online information of four countries, namely: Canada, The United Kingdom, Ireland, and the United States, and compared the readability of Website source origin and geographical origin. Only 17.2% of Webpages were found to be widely accessible. Vishala and Dexter [6] assessed the readability of COVID-19-related online content given by the government and public health agencies and departments. They discovered that all 18 Websites under study were easy to understand and had the reading level of grades 6 through 8.

3 Methodology

In this study, the readability of COVID-19-related Websites in India was evaluated. The Website or Webpage was searched by Google using the keywords: like "COVID-19-related Websites in India" or "COVID-19 information in India" or "COVID-19 resources in India." A total of 38 Websites belonging to various cities, states, and

centers were evaluated. The choice was toward selecting the public health Websites of the COVID-19 pandemic as of mid-May 2021. The evaluation was conducted on the English language of the homepage of every Website. If the Website was in any other language, it was translated to English before applying the test.

COVID-19 is a pandemic that has spread over both rural and urban areas. Therefore, it was required that the information be disseminated quickly to all the people, so it was essential to evaluate the reading ease of the text written on the Website. Readability is a metric for determining how easy or difficult a text is to read and comprehend. Since human evaluation is prone to errors and inconsistencies, so to ensure consistency and avoid errors caused by humanly limitations, the readability scores were computed using an online readability tool. There are various online tools available for computing readability. The online readability calculator [7] was used to calculate the readability of Websites to avoid human evaluation errors. The tool takes URL as an input, or plain text can also be copied and pasted into the tool to evaluate the readability score. In this paper, the URL of the Website under study is given as input to the online tool.

The various readability parameters, along with the definitions and formula, are taken from the WebFX tool as follows:

Flesch Kincaid Reading Ease (FKRE) test measures how easy or difficult the text is to understand. The measure is based on a score between zero and a hundred. A lower score indicates that the text is difficult to understand, while a higher score indicates the ease of understanding the text. The measure uses formula 1 to calculate the reading score.

$$\text{FKRE score} = 206.835 - 1.015 \times \left(\frac{\text{Words}}{\text{Sentences}} \right) - 84.6 \times \left(\frac{\text{Syllable}}{\text{Words}} \right) \quad (1)$$

Flesch Kincaid Grade Level (FKGL) test: The Flesch Kincaid Grade Level is used to assess a text's approximate reading grade level. The score obtained indicates the grade level. For example, if the grade number is six, an average student in the sixth can read and understand the content with ease. It is calculated using the formula 2:

$$\text{FKGL Score} = 0.39 \times \left(\frac{\text{Total Words}}{\text{Total Sentences}} \right) + 1.8 \times \left(\frac{\text{Total Syllables}}{\text{Total Words}} \right) - 15.59 \quad (2)$$

Gunning Fog (G.F.) score: The Gunning Fog Index is a metric that determines the number of years of education required to comprehend a passage of text on first reading. The Gunning Fog algorithm yields a grade level score ranging from 0 to 20. The grade level is determined by the score received. A score of 6 indicates that the text is appropriate for a 6th grader. It is a weighted average of how many words per phrase and how many lengthy words per word are there in the sentence. It is computed using formula 3:

$$\text{Gunning Fog Score} = 0.4 \times \left[\left(\frac{\text{Total words}}{\text{Total Sentences}} \right) + \left(\frac{\text{Complex Words}}{\text{Total Words}} \right) \right] \quad (3)$$

Coleman Liau Index (CLI) is another index of readability measured using the average number of characters and sentences per 100 words. The grade level is the floor and ceiling values of the score obtained. For example, if the score is 9.3, it is understood by 9th and 10th students. It is mainly used to assess educational text material but can be used for other text too. It is calculated using formula 4:

$$\text{Coleman Liau Index(CLI)} = 5.89 \times (\frac{\text{characters}}{\text{words}}) - 0.3 * (\frac{\text{sentences}}{\text{words}}) - 15.8 \quad (4)$$

Automated Readability Index (ARI): The automated readability index is a readability test to measure how easy it is to comprehend the text. A score of 1 means it can be understood by Kindergarten grade students of age 5–6 years old while a score. Similarly, a score of 12 means 12th-grade students can understand it, while the score of 18–24 can easily be understandable by the college student. The professor-level people can comprehend a score greater than 24. It is calculated using formula 5:

$$\text{Automated Readability Index (ARI)} = 4.71$$
$$\times \left(\frac{\text{characters}}{\text{words}} \right) + 0.5 \times \left(\frac{\text{words}}{\text{sentences}} \right) - 21.43 \quad (5)$$

SMOG Index: SMOG, a readability index, stands for "Simple Measure of Gobbledygook," which estimates how many years of education the average person needs to understand a text. It is calculated using the formula 6.

$$\text{SMOG Index} = 1.0430 \times (\sqrt{30 \times \left(\frac{\text{complex words}}{\text{sentences}} \right)}) + 3.1291 \quad (6)$$

4 Results

The results indicated that 52% of the Websites fell under the easy to read category, while 48% of Websites had a reading level of high school and above. The mean readability score for FKRE, FKGL, G.F., SMOG, CLI, ARI was 58.97, 7.04, 6.9, 7.25, 8.9, 2.9, respectively. Table 1 shows the statistical results of readability scores of the Website under study. Figure 1 shows the score of the various readability indices applied on the Website under study. The mean FKRE score of 58.97 indicates a "fairly difficult to read" score and is readable by 10th -11th-grade students. The highest FKRE score was reported was 101 by the Website https://covidinfo.rajasthan.gov.in. A score

Table1 Descriptive statistics of the readability indices of the websites

	N	Minimum	Maximum	Mean	Std. deviation
Flesch kincaid reading ease	38	6.50	101.50	58.97	21.37
Flesch kincaid grade level	38	0.70	31.60	7.04	4.83
Gunning fog score	38	2.80	31.80	6.91	4.85
SMOG index	38	3.40	18.80	7.26	2.68
Coleman liau index	38	−9.50	18.90	8.97	5.87
Automated readability index	38	−13.10	27.60	2.99	6.29

of 100 or more means the text is very easy to read and can be understood by 11 year old.

The public health professionals recommend a 6th-grade reading level to be a universal reading level [8]. Table 2 shows the result of FKRE score on Websites under study. The lowest FKRE score reported was 6.5 by https://covid19.assam.gov.in/ Website, which is interpreted as very difficult to read and can be understood only by college graduates. The mean readability was found to be of 6th and 7th g**rade.**

But another issue in India regarding online content dissemination is the information should be presented in native language. Table 3 shows the result of presence of multi-lingual option on Websi**te.**

In one of the studies, it was found that Website languages may severely limit Website access by users who are not provided with information in their native language [9]**.**

In a country like India, where there are 33 different languages and 2000 dialects [10], out of 38 Websites that were evaluated, 19 Websites were only in English while 15 had multilingual support, 3 Websites were in Hindi as well as in English while one Website was in English and a regional language. Figure 2 shows the result of readability indices of Websites under study.

5 Threat to Validity

Many of these tests are based on English language text, and the result may not be significant if the text is in a language other than English. Also, the scores obtained indicate the U.S. school level required by a person to understand the text. How much valid they are in the Indian context is another question to ponder. Also, different readability index gives different score and grade for the same text. Additionally, readability does not consider the meaning and context of the text. Low literacy levels in India are another barrier to online health knowledge.

Fig. 1 Bar chart showing representation of readability indices

6 Conclusion

A large population of India uses the Internet for healthcare information. This is due to the digital revolution, which has made the Internet accessible even in remote areas. There is a vast amount of COVID-19 content available online. In the study, it was found that the readability of content addressing COVID-19 is satisfactory. But language forms another barrier in information dissemination. Research is needed to

Table 2 Number of websites falling in different categories according to FKRE

Score	School level	Interpretation	No. of websites
90.0–100.0	5th grade	Very easy to read	2
80.0–90.0	6th grade	Easy to read	0
70.0–80.0	7th grade	Fairly easy to read	13
60.0–70.0	8th and 9th grade	Plain english	5
50.0–60.0	10th to 12th grade	Fairly difficult to read	5
30.0–50.0	College	Difficult to read	9
0.0–30.0	College graduate	Very difficult to read	4

Table 3 Number of websites in different language

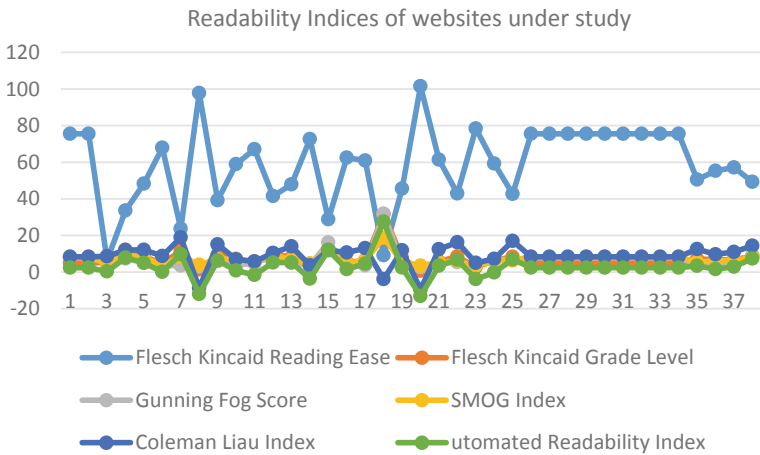| Total sites	38
Only English	19
Hindi/English	3
Multilingual	15
Regional	1

Fig. 2 Readability indices of websites

make the readability measures adapt to multilingual pages focusing on context and its meaning in the language.

The pandemic nature of COVID-19 requires adequate and effective communication and dissemination of reliable information from the healthcare people to the general public. This can be done by creating online material written at the recommended reading level and comprehensible to help the public take informed decisions about risk mitigation measures.

References

1. Grose EM, Holmes CP, Aravinthan KA, WV ,Lee JM (2021) Readability and quality assessment of internet-based patient education materials related to nasal septoplasty. J Otolaryngol—Head Neck Surg 50(16)
2. Kher A, Johnson S, Griffith R (2017) Readability assessment of online patient education material on congestive heart failure. Adv Prev Med 2017:9780317. https://doi.org/10.1155/2017/9780317
3. Szmuda T, Özdemir C, Ali S, Singh A, Syed MT, Słoniewski P (2020) Readability of online patient education material for the novel coronavirus disease (COVID-19): a cross-sectional health literacy study. Public Health 185:21–25
4. Jessica K, Paloma T, Tayler BB, Erica JT, Charlesnika TE, William AG, Emily SM, Elizabeth MSL(2021) Readability, content, and quality of COVID-19 patient education materials from academic medical centers in the United States. American J Infect Control 49(6):690–693. ISSN 0196–6553. https://doi.org/10.1016/j.ajic.2020.11.023
5. Worrall AP, Connolly MJ (2020) O'Neill A (2020) Readability of online COVID-19 health information: a comparison between four English speaking countries. BMC Public Health 20:1635. https://doi.org/10.1186/s12889-020-09710-5
6. Mishra V, Dexter JP (2020) Comparison of readability of official public health information about COVID-19 on websites of international agencies and the governments of 15 Countries. JAMA Netw Open. 3(8):e2018033.https://doi.org/10.1001/jamanetworkopen.2020.18033
7. Webpagefx. https://www.webfx.com/tools/read-able/. Accessed May'2021
8. McKenzie JF, Neiger B, Thackeray R (2017) Planning, implementing, and evaluating health promotion programs: a primer, 7th edn. Pearson, New York
9. Kralisch A, Berendt B (2005) Language-sensitive search behaviour and the role of domain knowledge. New Rev Hypermedia Multimedia 11:221–246
10. Facts about India-Indian languages. http://www.webindia123.com/india/people/language.htm