

# Readability Ease of Online Hearing-Related Information in Hindi

Seema Diwan, Rebecca J. Kelly Campbell<sup>1</sup>

Audiologist, Hearing Life, Tauranga, <sup>1</sup>Department of Communication Disorders, University of Canterbury, Christchurch, New Zealand

## Abstract

**Introduction:** The purpose of this study was to assess the readability of hearing-related Internet information in Hindi language. **Methods:** Native Hindi speakers identified five Hindi keywords relating to hearing problems that were used to search for hearing-related web pages. These key terms were entered one by one into Google Bharat, the Hindi version of Google India. The uniform resource locators were recorded for the first ten web pages resulting from that search. Each web page was assessed according to the inclusion and exclusion criteria. The reading grade level (RGL) for the resulting 25 web pages were analyzed using Readability Hindi 1 (RH1) and Readability Hindi 2 (RH2) formulas. The paragraphs with lowest and highest RGL were identified and used for a cloze test. Ten participants were recruited after applying inclusion and exclusion criteria. They were instructed to complete the cloze test. **Results:** The mean RGL of hearing-related web pages published in Hindi was not significantly different from the recommended value. The mean RGL calculated by RH1 was significantly higher than the mean RGL calculated by RH2; however, there was a significant and positive correlation between the RGL values calculated by RH1 and RH2. No significant differences in cloze scores were found between the paragraph with the highest RGL and the paragraph with the lowest RGL. **Conclusions:** The RGL calculated by the formulas was within the recommended value, which indicates the hearing-related material available on the Internet in Hindi is easy to read. However, the results of readability ease calculated by the cloze test suggested that the paragraphs with maximum RGL and minimum RGL were not significantly different from each other.

**Keywords:** Cloze test, hearing loss, readability

## INTRODUCTION

Readability comprises many elements that can affect the ability of readers to understand the material and read it at an optimal pace.<sup>[1]</sup> In other words, readability is a factor that makes the text easy to read and understand.<sup>[2]</sup> Reading grade level (RGL) is a number assigned to the level of complexity of the text.<sup>[3]</sup> For example, if the RGL is five, the text to read is suitable for a 5<sup>th</sup> standard school student.<sup>[4,5]</sup> Researchers have recommended that health-related material should be written at an RGL of 5 or 6. If the RGL is between 6 and 8, the material can be considered adequate. If the RGL is above 9, the material can be considered unsuitable.<sup>[6]</sup> Readability formulas in Hindi were developed by researchers in a series of experiments to identify the salient features of Hindi-affecting readability.<sup>[6]</sup> Their study helped in the development of two new readability measures or formulas: Readability Hindi 1 (RH1) and Readability Hindi 2 (RH2) which are measured in RGL.<sup>[6]</sup>

Readability calculated by formulas cannot always accurately predict comprehension and some difficulties were found by researchers.<sup>[6]</sup> Researchers also<sup>[7]</sup> argued that words *per se*

are not the measure of reading difficulty, but rather it is their relationship with each other and their comprehension. The cloze test<sup>[7]</sup> was published to measure a person's understanding of a text. The inherent theory of this test is the ability of a person to fill in the missing words of a text, based on the surrounding context, thereby providing closure to the meaning of the text. It is hard to complete more than 65% of the deleted words even by readers with advanced reading skills.<sup>[2]</sup> Assisted reading texts need a cloze score of 35% or more. A high cloze score and low RGL are needed for unassisted reading text.

Another factor affecting the ability of Internet users to understand online health information depends, in part, on their health literacy.<sup>[8]</sup> Health literacy further depends on the reading ability of the person.<sup>[9]</sup> People with a low level of reading comprehension are 1.5–3 times more prone to poor health outcomes than people who read at a higher comprehension level.<sup>[9]</sup>

**Address for correspondence:** Seema Diwan,  
No. 6/37, Mclean Street, Tauranga 3110, New Zealand.  
E-mail: whitesdr@gmail.com

### Access this article online

#### Quick Response Code:



**Website:**  
www.jisha.org

**DOI:**  
10.4103/jisha.JISHA\_20\_18

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

**For reprints contact:** reprints@medknow.com

**How to cite this article:** Diwan S, Campbell RJ. Readability ease of online hearing-related information in Hindi. J Indian Speech Language Hearing Assoc 2018;32:62-6.

According to a report,<sup>[10]</sup> “India has overtaken the United States to become the second largest Internet market, with 333 million users, trailing China’s 721 million (para 1).” Increasing Internet usage in India requires more research in regard to readability ease of the online text available which can directly affect health outcomes.<sup>[9]</sup> To date, there is no research available that has investigated the readability of hearing-related information available on the Internet in the Hindi language. Given the large population of Hindi speakers in the world and the increasing usage of the Internet, there is a clear demand for the assessment of text readability available on Hindi language web pages.

### Aims and hypothesis

The aims of this study were to (1) examine the mean readability (RGL) of online hearing-related web pages in Hindi, (2) examine the relationship between two readability formulas: RH1 and RH2 in Hindi, and (3) examine the relationship between RGL (RH1 and RH2) and comprehension (cloze score) of the web pages.

The planned alternative (research) hypotheses were as follows:

- Hypothesis 1: The mean readability score (RGL) of hearing-related web pages published in Hindi is significantly higher than 6
- Hypothesis 2: There is a significant difference between the readability score (RGL) calculated by RH1 and RH2 formulas
- Hypothesis 3: There is a significant correlation between the readability scores (RGL) calculated by RH1 and RH2
- Hypothesis 4: There is a significant difference between the cloze scores and RGL of the web pages.

## METHODS

This study was conducted at University of Canterbury as a requirement of thesis for Master of Audiology program in the Department of Communication Disorders. The study methodology was reviewed and approved by the University of Canterbury Human Ethics Committee.

### Identification of search terms

The identification of keywords for the Internet search was done by recruiting a group of people who spoke Hindi as their first language. The 25 informants were identified through Facebook and through personal links. They were asked the following question in Hindi: “If you realize that you are having hearing problems or difficulties and you want to look for general information about a hearing problem and its treatment, which search terms will you try on the Google search engine? Please do not hesitate to mention as many as you like.” Two individuals declined to identify search terms because they said they did not use Hindi language for Internet searches. The search phrases or keywords identified by the informants were as follows: Kaan ki samasya (ear problems), Kaan ki masheen (hearing aid), Kaan ki pareshani (ear troubles), Kaan mein dard (ear pain), Sunne mein pareshani (hearing trouble), Kam sunai dena (hearing impairment), and Baherepan key lakshan (signs of deafness). Out of these seven phrases, five search phrases were selected

on the basis of their use by two or more informants; these were: Sunne mein pareshani (hearing trouble), Kaan ki samasya (ear problems), Kaan ki masheen (hearing aid), Baherepan key lakshan (signs of deafness), and Kam sunai dena (hearing impairment).

### Internet search

A 13-inch MacBook Pro having OS X El Capitan (Apple Inc., Cupertino, California, USA) operating system version 10.11.6 was used to perform the online search. Each search phrase was entered one by one into Google India (google.co.in) Hindi version (named Google Bharat). The uniform resource locators of the first ten web page results obtained after entering each search phrase were recorded. The web pages were used as units for analysis, so one web page was treated as one unit or participant.

### Passage selection

The inclusion criteria for the selection of web pages were as follows: (1) must be in Hindi, (2) must contain hearing or hearing impairment-related information, (3) must be available to the public, and (4) must contain information about the organization hosting the web page. There was no information about the organization hosting the web page on some web pages. On those web pages, additional research was conducted to obtain information about the organizations and the information was collected by searching through a separate Google search.

The exclusion criteria (chosen on the basis of previous research by Hsu<sup>[11]</sup>) for the selection of web pages were included: (1) a Google-identified advertisement; (2) a video; (3) a directory listing; (4) a web page containing < 100 words; (5) a web page containing information on tinnitus, otitis media, tumors, and vestibular disorders; and (6) a web page containing images only.

The relevant content in Hindi on each web page was copied and pasted into a Microsoft Word document. Each sentence was numbered and a random generator in an Excel Spreadsheet was used to select the first sentence for the analysis. The portion of the text to be analyzed was composed of the first sentence starting from that random number and the next 100 words/characters, confirming that the last sentence was a full sentence. In this manner, each selected paragraph consisted of at least 100 words or characters.

### Readability Hindi 1 and 2 formulas

Each selected paragraph was entered in a computational tool<sup>[6]</sup> and RH1 and RH2 formulas were used to calculate the readability of the paragraph.

$$RH1 = -2.34 + 2.14 \times AWL + 0.01 \times PSW$$

$$RH2 = 0.211 + 1.37 \times AWL + 0.005 \times JUK$$

AWL = Average number of syllables per word

PSW = Polysyllabic words

JUK = Jukta Akshars (consonants together in clusters).

The results obtained by this analysis of the computational tool were recorded in an Excel Spreadsheet. Mean RGL of all the selected paragraphs was calculated by taking the arithmetic average of the RGL obtained by RH1 and RH2.

### Cloze test

The paragraphs on the web pages that resulted in the lowest (easiest to read) and the highest (most difficult to read) RGLs were identified and used to expand the extent of understanding between readability ease and comprehension. In the identified paragraphs, every fifth word was replaced by a blank.

### Participants

Ten randomly selected participants living in the campus of University of Canterbury were recruited over a 2-week period according to the inclusion and exclusion criteria detailed in Table 1. The exclusion criteria attempted to ensure that the knowledge about hearing health would not influence the responses given by participants.

The inclusion criteria for the selection of the participants were as follows: (1) participants must be aged 18 years and above and (2) participants must be native Hindi speakers of any gender and the exclusion criteria were as follows: participants must not have any expertise in the hearing health industry.

### Procedure

Participants were identified from a list of Facebook friends. Participants were instructed to sign the consent form after reading the information sheet fully and to fill in the demographic questionnaire. Two paragraphs in Hindi (selected after readability analysis) having blanks (every fifth word) were displayed on this sheet. The participant needed to fill in the blanks, using only the information provided in the paragraph. If the task was found difficult, participants were encouraged to make a guess. All the participants returned the forms and none of them withdrew from the study.

## RESULTS

### Aim 1

Readability analyses were performed for 25 web pages (out of 50 web pages: 5 search terms + 10 results) which were obtained after removing duplicates and applying inclusion and exclusion criteria. The minimum RGL of hearing-related web pages published in Hindi obtained by RH1 was 3.03, and by RH2, it was 3.12. The maximum RGL by RH1 was 11.40 and by RH2 was 5.74. Descriptive statistics obtained from RH1 and RH2, and the mean RGL of the web pages are illustrated

**Table 1: Inclusion criteria and exclusion criteria for participants**

Inclusion Criteria	Exclusion Criteria
1. Participants must be aged 18 years and over	Participants must not have any expertise in the hearing health industry
2. Participants must be native Hindi speakers of any gender	

in Figure 1. The data did not meet the assumption of normality; therefore, nonparametric statistics were used to test the study hypotheses [Figure 1].

It was hypothesized that the mean RGL of hearing-related web pages published in Hindi is significantly different from 6. A one-sample Kolmogorov–Smirnov test showed that the mean RGL of hearing-related web pages published in Hindi is 5.33, which is not significantly different from the recommended value of 6 ( $P = 0.200$ ).

### Aim 2

A Wilcoxon signed-rank test showed that the mean RGL calculated by RH1 was significantly higher than the mean RGL calculated by RH2 ( $Z = -4.157$ ,  $P < 0.001$ ). Spearman's rho ( $r_s$ ) revealed that there was a significant and positive correlation between the RGL values calculated by RH1 and RH2 ( $r_s = 0.954$ ,  $P < 0.001$ ).

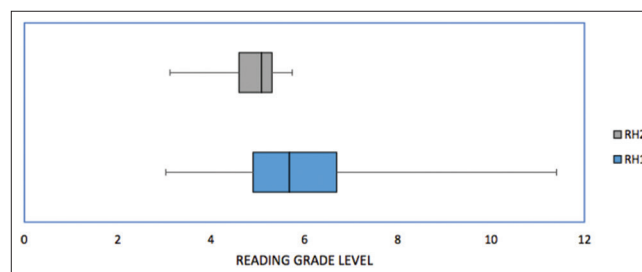
### Aim 3

Table 2 shows the cloze scores of the 10 participants (mean age: 25; average years of education: 18) from the two paragraphs. A Wilcoxon signed-rank test indicated that there was no significant difference between the cloze scores of the web page with the lowest and highest mean RGL ( $Z = -1.779$ ,  $P = 0.075$ ) with an effect size = 0.40 (Cohen's d).

## DISCUSSION

The mean RGL of online hearing-related information in Hindi did not exceed the recommended level for written health information. In this study, we applied the recommended RGL of 6 which exists in research performed for English health information.<sup>[4,5]</sup> The mean RGL of the online hearing-related web pages in Hindi was 5.33. In particular, the RH1 formula had a mean of 5.95 and RH2 a mean of 4.72, which suggest that Hindi online hearing-related information is not hard to read. These values indicate that the Hindi text available on these web pages is easy to read and there is no need to rewrite it in a simple language. Hence, the first research hypothesis is not supported.

This is the first study conducted to identify RGL of online hearing-related information in Hindi. However, similar



**Figure 1:** Box plot of reading grade level of Hindi web pages. The boxes represent the middle 50% of the reading grade levels, the vertical line represents the median reading grade level and the whiskers represent the minimum and maximum reading grade levels. RH1 = Readability Hindi 1 and RH2 = Readability Hindi 2

**Table 2: Cloze score data obtained from participants and their descriptive summary**

Participants	Cloze score on paragraph with maximum mean RGL (%)	Cloze score on paragraph with minimum mean RGL (%)
1	29.16	34.7
2	37.50	39.13
3	41.60	39.13
4	50.00	41.60
5	50.00	41.66
6	54.16	43.74
7	58.30	45.83
8	58.33	47.80
9	58.33	50.00
10	66.66	50.00
Mean	50.40	43.3
SD	11.36	5.06

RGL: Reading grade level; SD: Standard deviation

types of studies conducted to find the RGL in English and Chinese did not show the same results. In a study completed by Hsu,<sup>[11]</sup> the mean RGL of Chinese web pages was 7.32 with a range of 4.16–12.25. In a systematic review,<sup>[9]</sup> authors demonstrated that the mean RGL of online health-related information of web pages in English ranged from 9 to over 14. The researchers concluded that there is enough evidence to say that the hearing-related information available on web pages in English has poor readability and this issue must be addressed immediately to provide maximum benefits to consumers from that information. However, our study did not achieve the same results and it was designed on the basis of the previous studies to remove biases.<sup>[12]</sup> A possible explanation is that they are easy to read, or this might be that the Hindi readability formulas lack special lexical attributes of text used in hearing-related information.

It was hypothesized in the second hypothesis that there is a significant difference in readability score obtained by RH1 and RH2 readability formulas. The results supported this research hypothesis because readability scores obtained by RH1 were significantly higher than those obtained by RH2. There is no literature to support or contradict these findings. However, a possible explanation could be in the structure of their formulas: RH1 ( $-2.34 + 2.14 \times \text{AWL} + 0.01 \text{ PSW}$ ) and RH2 ( $0.211 + 1.37 \times \text{AWL} + 0.005 \times \text{JUK}$ ) were designed by Sinha and Sharma.<sup>[6]</sup> The researchers validated these two formulas but did not mention the type of text used by them while validating. Sinha and Sharma<sup>[6]</sup> also observed that AWL, JUK, PSW, and PSW30 are key features contributing toward readability in Hindi, and in the current study, we found a significant difference between RGL obtained by RH1 and RH2.

The next hypothesis was to find if there is a significant correlation between the readability scores calculated by RH1 and RH2. The results indicated a positive significant correlation between the readability scores of the RH1 and RH2 formulas.

Due to the unavailability of research on the relationship between these two readability formulas, we cannot comment on the consistency or contradiction of the present findings. However, the results explain that either of these formulas can be used to calculate the readability score of a Hindi text, due to the strong, positive correlation between them, supporting the hypothesis.

The statistical analysis did not support hypothesis 4, revealing no significant difference between the cloze scores and RGLs of the web pages. The mean cloze score of the text with maximum mean RGL and with minimum mean RGL (both obtained by readability formulas) were 50% and 43.3%, respectively, which suggests that the online hearing material available in Hindi text can be used but may require some additional guidance or teaching.<sup>[4]</sup> It suggests that the paragraph with highest mean RGL is as understandable as the paragraph with lowest RGL, which is not a finding we expected. Moreover, if we look at the effect size ( $d = 0.40$ ), it shows a difference between the findings of two paragraphs. However, the lack of statistical significance could be due to the small sample size and the absence of normal distribution of the data.

Another possible reason behind these findings is that the validated readability formulas need further evaluation, and hence, it is necessary to identify if they can be applied to a hearing-related text in Hindi. It may be possible that the participants of this study are not representative of consumers of online hearing information in Hindi because they were well educated and fairly young. No research was found in the literature on the question of readability of online hearing-related information in Hindi obtained by a cloze test.

In this study, efforts were made to replicate the search strategy that had been used in other readability studies performed in different languages<sup>[11]</sup> so that results could be compared. However, it is possible that different search key terms in a different style can be used by people speaking different languages. This can further impact the results of the web pages obtained by the search using specific key terms. Another limitation in the search strategy was that the search terms were decided upon by putting up a question on the Facebook friend list because it could represent people belonging to a particular demographic group. In addition, it is not necessary that each person in the target population uses Facebook. In India, only 15% of the whole population use Facebook;<sup>[13]</sup> therefore, the people recruited by Facebook cannot represent the whole Indian population.

Another limitation is the low number of hearing web pages available in Hindi compared to those available in English. Around 80% of Web information was available in English until 1990, and by 2011, the information extended to Chinese, French, German, Russian, and Spanish, but Indian languages still lagged behind.<sup>[14]</sup> The idea of translating medical information in Hindi through Google search engine was introduced recently,<sup>[15]</sup> which can explain the low number of online Hindi hearing-related web pages. In addition, when



searching for web pages in Hindi, most of the web pages used English terminology but written in Hindi alphabets and two web pages were completely written in Hindi alphabets with English terminology, these were excluded from the study. Selecting the content in Hindi caused difficulties in identifying the readability of that web page; therefore, the sample size of the study was small and a cloze test was used to get more information about the readability ease of Hindi text related to hearing available on the Internet. Due to the unavailability of readability research in Hindi, it was difficult to compare the results of this study with others. There is no literature available that discusses the validity of the readability formulas in Hindi. Sinha and Sharma<sup>[6]</sup> designed these formulas and validated them. No other researcher has validated these formulas.

This study is the first step toward the readability analysis of hearing-related information available on the Internet in Hindi. Because the readability formulas in Hindi were not designed for health-related web pages, this study was not able to practically assess the readability grade level of hearing-related web pages. Therefore, future research is required to update the present readability formulas so that health information can also be assessed at the level of reading difficulty. Internet search in Hindi is becoming popular among the Indian population and web pages containing health-related content in Hindi are also increasing day by day,<sup>[15,16]</sup> which clearly indicates the need for more research.

## CONCLUSIONS

This study identified the readability ease of online hearing-related information available in Hindi available to consumers who speak Hindi as their first language. Readability was analyzed using the RH1 and RH2 readability formulas proposed by Sinha and Sharma<sup>[6]</sup> and a cloze test. The results of the study demonstrated that RGL calculated by the formulas was within the recommended value, which means the hearing-related material available on the Internet in Hindi is easy to read. However, the results of readability ease calculated by the cloze test suggested that the paragraphs with maximum RGL and minimum RGL were not significantly different from each other in their level of difficulty in understanding.

From a clinical perspective, it means that clinicians should be careful before recommending any online hearing material to their patients based on the RGL. Moreover, readability formulas should be evaluated further for the specific user

population and the content of the information provided on the Internet.

## Financial support and sponsorship

UC library.

## Conflicts of interest

There are no conflicts of interest.

## REFERENCES

1. Dale E, Chall JS. The concept of readability. *Elem Engl* 1949;26:19-26.
2. DuBay WH. The Principles of Readability. Costa Mesa, CA: Impact Information; 2004. p. 70.
3. The Law Dictionary, Featuring Black's Law Dictionary Free Online Legal Dictionary; 1910. Available from: <http://www.thelawdictionary.org/reading-grade-level-rgl/>. [Last accessed on 2017 Oct 16].
4. Doak CC, Doak LG, Root JH. Teaching Patients with Low Literacy Skills. 2nd ed. Philadelphia, PA: J. B. Lippincott Company; 1996. p. 212.
5. Weiss BD; National Work Group on Literacy and Health. Communicating with patients who have limited literacy skills: Report of the national work group on literacy and health. *J Fam Pract* 1998;46:168-76.
6. Sinha M, Sharma S, Dasgupta T, Basu A. New Readability Measures for Bangla and Hindi Texts; 2012. Available from: <http://www.aclweb.org/anthology/C12-2111>. [Last accessed on 2017 Sep 05].
7. Taylor WL. "Cloze procedure": A new tool for measuring readability. *Journal Bull* 1953;30:415-33.
8. Sorensen K, Van den Broucke S, Fullam J, Doyle G, Pelikan J, Slonska Z, *et al*. Health literacy and public health: A systematic review and integration of definitions and models. *BMC Public Health* 2012;12:80.
9. Dewalt DA, Berkman ND, Sheridan S, Lohr KN, Pignone MP. Literacy and health outcomes: A systematic review of the literature. *J Gen Intern Med* 2004;19:1228-39.
10. Union IT. Committed to Connecting the World Geneva; 2016. Available from: <http://www.itu.int/en/mediacentre/Pages/2016-PR35.aspx>. [Last updated on 2016 Sep 15; Last accessed on 2017 Nov 19].
11. Hsu PH. Readability of Hearing Related Internet Information in Traditional Chinese. Christchurch: University of Canterbury; 2017.
12. Laplante-Lévesque A, Thorén ES. Readability of internet information on hearing: Systematic literature review. *Am J Audiol* 2015;24:284-8.
13. Statista. Number of Social Network Users in India from 2015 to 2022 (in Millions) Hamburg, Germany; 2017. Available from: <https://www.statista.com/statistics/278407/number-of-social-network-users-in-india/>. [Last accessed on 2017 Sep 11].
14. Bhattacharya A. India's Internet Users have More Faith in Content that's not in English USA2017. Available from: <https://www.qz.com/972844/indias-internet-users-have-more-faith-in-content-thats-not-in-english-study-says/>. [Last accessed on 2017 Nov 10].
15. Google Organization. Google 'Health Speaks' for Online Medical in for in Hindi, THE HINDU; 2016. Retrieved from: <https://www.thehindu.com/sci-tech/technology/Google-IsquoHealth-Speaksrsquo-for-online-medical-info-in-Hindi/article15906333.ece>. [Last accessed on 2017 Nov 10].
16. Share of Internet Adoption Levels Among non-English Users Across India in 2016, by language; 2018. Retrieved from: <https://www.statista.com/statistics/718476/internet-adoption-levels-among-non-english-users-by-language-india/>. [Last accessed on 2017 Nov 10].