

Assessing Readability of Skin Cancer Screening Resources: A Comparison of Online Websites and ChatGPT Responses

Elissa J. Goorman, BA^{1,2}; Sukul Mittal, BS^{1,3}; Jennifer N. Choi, MD¹

¹Department of Dermatology, Northwestern University Feinberg School of Medicine, Chicago, IL; ²Jacobs School of Medicine and Biomedical Sciences, Buffalo, NY; ³University of Texas Southwestern School of Medicine, Dallas, TX

Background

- Skin cancer is a significant public health concern
- Lack of official guidelines leads patients to turn to online resources for guidance on screening recommendations
- Readability of online patient education materials is often too complex for the public
- Common readability tests include
 - Flesch-Kincaid Reading Ease (FRE)
 - Flesch-Kincaid Grade Level (FKGL): Most used; can underestimate difficulty
 - Simple Measure of Gobbledygook (SMOG) Index: Gold standard
 - Gunning Fog Index (GFI)
 - Coleman-Liau Index (CLI)
- ChatGPT offers access to interactive, conversational responses tailored to individual queries, potentially improving health information accessibility
- Readability and understandability of both ChatGPT's responses and traditional online sources must be evaluated to ensure that comprehensible information is available

Objective

- To assess and compare the readability of websites and ChatGPT-generated responses related to skin cancer screening, offer insights into the accessibility of these resources, and identify potential areas for improvement

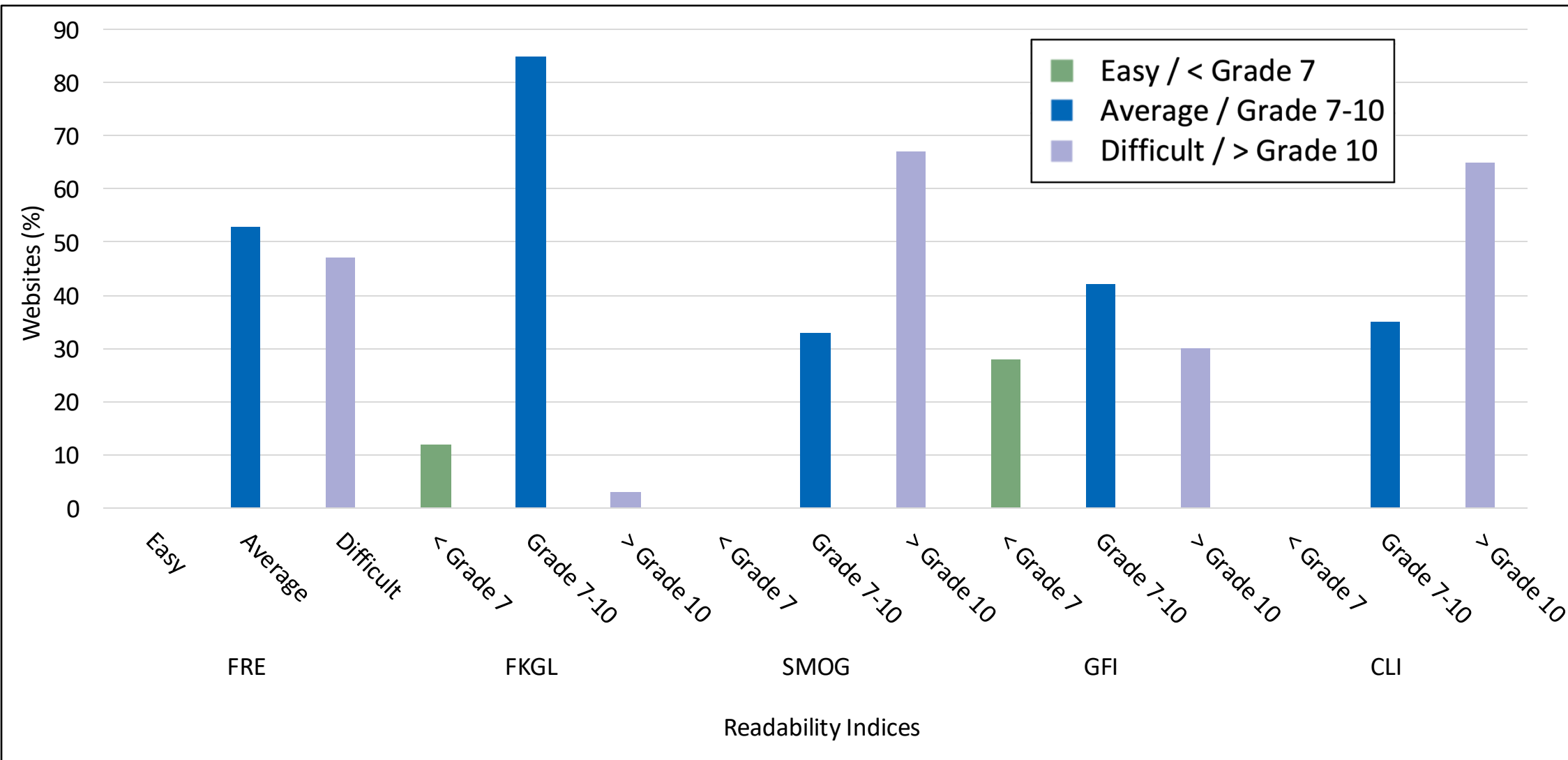
Methods

- “Skin cancer screening” was searched on Chrome Incognito browser, and the first 100 websites were reviewed
- Websites were categorized as university-hosted or non-university-hosted
- Five questions were submitted to ChatGPT in three separate Incognito sessions each, to account for response variability
 1. What should I expect at a skin cancer screening?
 2. When should I get my skin cancer screening?
 3. Where should I get my skin cancer screening?
 4. Who should do my skin cancer screening?
 5. How do I do a skin cancer screening at home?
- Responses were analyzed alongside website content using Readable.io, a validated tool for automating readability assessments
- The metrics applied included the FRE, FKGL, SMOG Index, GFI, and CLI
- Data analysis was conducted using SPSS version 29.0. Continuous variables were described using means and standard deviations. Student’s t-tests were used to determine significance, which was set at a threshold of $p < 0.05$.

Table 1. Overview of readability and understandability scores used for all websites, university-hosted websites, non-university hosted websites, and ChatGPT responses

Responses	Average Scores (SD)	Corresponding Reading Grade Level	Corresponding Readability Difficulty
All Websites (n = 60)			
Flesch-Kincaid Reading Ease	58.47 (10.69)	Late high school	Moderately difficult
Flesch-Kincaid Grade Level	8.16 (1.82)	Middle school	Average
SMOG Index	10.72 (1.46)	Late high school	Difficult
Gunning Fog Index	9.56 (2.11)	Early high school	Moderately difficult
Coleman-Liau Index	10.16 (2.00)	Late high school	Difficult
University Hosted Websites (n = 16)			
Flesch-Kincaid Reading Ease	58.8 (8.69)	Late high school	Moderately difficult
Flesch-Kincaid Grade Level	7.79 (1.62)	Middle school	Average
SMOG Index	10.51 (1.24)	Late high school	Difficult
Gunning Fog Index	9.07 (1.83)	Early high school	Moderately difficult
Coleman-Liau Index	10.31 (1.77)	Late high school	Difficult
Non-University Hosted Websites (n = 44)			
Flesch-Kincaid Reading Ease	58.35 (11.41)	Late high school	Moderately difficult
Flesch-Kincaid Grade Level	8.29 (1.88)	Middle school	Average
SMOG Index	10.79 (1.54)	Late high school	Difficult
Gunning Fog Index	9.74 (2.19)	Early high school	Moderately difficult
Coleman-Liau Index	10.11 (2.10)	Late high school	Difficult
ChatGPT Responses			
Flesch-Kincaid Reading Ease	46.27 (11.44)	College	Difficult
Flesch-Kincaid Grade Level	11.05 (2.66)	Early high school	Difficult
SMOG Index	13.33 (2.29)	College	Difficult
Gunning Fog Index	13.39 (3.74)	College	Difficult
Coleman-Liau Index	11.79 (1.92)	Late high school	Difficult

Figure 1. Distribution of website readability scores by category



Results

- Of the 60 websites that met inclusion criteria, the average SMOG Index score was 10.16, indicating a late high school reading level (Table 1)
- Average SMOG Index score for ChatGPT responses was 13.33, corresponding to a college reading level (Table 1)
- 16 (26.7%) of the websites were university hosted, while 44 (73.3%) were not
- There were no significant differences between university and non-university websites across all readability metrics (Table 1)
- 47% of websites had FRE scores in the difficult range (0–59), while 53% scored in the average range (60–79) and none were categorized as easy (80–100) (Fig. 1)
- On the FKGL, 88% of websites were written above a 6th-grade level, with two websites also exceeding the 10th-grade level (Fig. 1)
- SMOG Index and CLI show most websites required a 10th-grade reading level (Fig. 1)
- ChatGPT responses had higher average difficulty scores than websites, a finding that was significant across all readability indices (p-values ranging from 0.00021-0.0004)

Conclusion

- Traditional websites and AI-generated responses both exceed the recommended literacy level and present readability challenges that could hinder patient comprehension of skin cancer screening information
- There is a clear need to refine health communication strategies
- Future efforts should focus on creating patient-centered content that is not only informative but also accessible across varying literacy levels
- Limitations: choice of AI model, choice of readability indices, limit to first 100 sites

References

1. Force, U.P.S.T., Screening for Skin Cancer: US Preventive Services Task Force Recommendation Statement. JAMA, 2016. 316(4): p. 429-435.
2. Force, U.P.S.T., Screening for Skin Cancer: US Preventive Services Task Force Recommendation Statement. JAMA, 2023. 329(15): p. 1290-1295.
3. Chu, J.T., et al., How, When and Why People Seek Health Information Online: Qualitative Study in Hong Kong. Interact J Med Res, 2017. 6(2): p. e24.
4. Basch, C.H., et al., Readability of online material related to skin cancer. Public Health, 2018. 163: p. 137-140.
5. MacLean, S.A., et al., Readability of information on colonoscopy preparation on the internet. Health Promot Perspect, 2018. 8(2): p. 167-170.
6. Fitzsimmons, P.R., et al., A readability assessment of online Parkinson's disease information. J R Coll Physicians Edinb, 2010. 40(4): p. 292-6.
7. OpenAI. ChatGPT: Optimizing Language Models For Dialogue. 2022; Available from: <https://openai.com/blog/chatgpt/>
8. Health, N.I.o., How to Write Easy-to_Read Materials. Last accessed on, 2018.
9. Weiss, B.D., Health literacy. American Medical Association, 2003. 253: p. 358.
10. Kaundinya, T., S. El-Behaedi, and J.N. Choi, Readability of Online Patient Education Materials for Graft-Versus-Host Disease. J Cancer Educ, 2023. 38(4): p. 1363-1366.
11. Henrikson, N.B., et al., Skin Cancer Screening: Updated Evidence Report and Systematic Review for the US Preventive Services Task Force. JAMA, 2023. 329(15): p. 1296-1307.
12. Fitzpatrick, P.J., Improving health literacy using the power of digital communications to achieve better health outcomes for patients and practitioners. Front Digit Health, 2023. 5: p. 1264780.