

Online Health Education About Sialendoscopy: A Study on Readability and Reliability

OTO Open 2025, Vol. 9(2):e70117

© 2025 The Author(s). OTO Open published by Wiley Periodicals LLC on behalf of American Academy of Otolaryngology-Head and Neck Surgery Foundation.
DOI: 10.1002/oto2.70117
http://oto-open.org

Wiley

Rohan Singh, BA, MD (Student), Christopher Badger, MD, and Arjun S. Joshi, MD

Abstract

Objective. Sialendoscopy is a diagnostic and interventional treatment for patients with salivary disease. Patients and physicians leverage website information to acquire knowledge about sialendoscopy; thus, understanding the quality of this information is essential. This study analyzes the quality and readability of online information on sialendoscopy.

Methods. "Sialendoscopy" was searched on Google, and the first 100 websites were evaluated. Each website was required to meet three criteria for inclusion: accessible when opened, content deemed relevant, and available in written format. Four validated readability and two validated reliability metrics were utilized. Additionally, a separate analysis was conducted for the top nine websites in the search engine, given most web traffic occurs on Google's first page.

Results. In assessing readability, the mean Flesch Reading Ease Score for included and the top nine websites was 36.2 and 39.5, respectively, with a *P*-value of .543. Both scores aligned with the "difficult to read" category. Other readability metrics aligned with high school reading levels. For reliability, the mean Discern score for the included and the top nine websites was 36.9 and 45.0, respectively, with a *P*-value of .030. These scores aligned with the "poor" and "fair" categories, respectively.

Discussion. The low readability and reliability scores implied that the online health information on sialendoscopy is not easily understandable at a reading level appropriate for the general public. Our findings showed that the most readable and highest quality websites were not the highest ranked in our search results. Factors such as search engine algorithms and complex medical terminology used in these informative websites contribute to the lack of readability and relability of online health education.

Implications for Practice. As Al evolves, future studies should be conducted to assess its impact on readability and reliability of online health information. There is an opportunity to adjust search engine algorithms, collaborate with communications specialists, and utilize new technologies, such as artificial intelligence chatbots, for the benefit of health seekers

Keywords

artificial intelligence, online health information, readability, reliability, sialendoscopy

Received December 4, 2024; accepted April 4, 2025.

nline health information is important to patients. According to Pew Research Center, 81% of individuals "rely a lot" on their research before making big decisions. Roughly 35% of US adults have used the internet to evaluate what medical conditions they or others possess. Given the ubiquity of patients seeking health information, online content should be informative and understandable.

Sialendoscopy is a minimally invasive procedure used to diagnose and treat salivary gland pathologies, such as ductal stenosis and sialolithiasis.³ Sialendoscopy is utilized when other conservative and preventive approaches are unsuccessful.⁴ Given the complexity of this procedure, the need for high-quality, easily readable information is important for patients.

The American Medical Association (AMA) and the United States Department of Health and Human Services (USDHHS) state that information written above a fifth- or sixth-grade reading level needs to be simplified for the general public.⁵ The National Institute of Health (NIH) found that encountering difficult text early in a passage increases the odds that an individual stops reading before completing the passage.⁵ Standardized scoring systems have been developed to assess accuracy, clarity, relevancy, attribution, and currency of online treatment information.^{6,7}

Several studies have scrutinized the readability and quality of online health information, but limited research

Division of Otolaryngology-Head and Neck Surgery, George Washington University School of Medicine and Health Sciences, Washington, DC, USA

Corresponding Author:

Rohan Singh, BA, MD, Division of Otolaryngology—Head and Neck Surgery, George Washington University School of Medicine and Health Sciences, 2300 M Street, NW, 4th Floor — Otolaryngology, Washington, DC 20037, USA. Email: rohanameetsingh@gmail.com

has been conducted on otolaryngology procedures.^{5,8-12} Only the most common otolaryngology procedures have been analyzed for readability and quality.⁸⁻¹⁰ Our paper analyzes the quality and readability of online information available to patients seeking education on sialendoscopy.

Materials and Methods

This study was deemed exempt from the George Washington University (GW) Institutional Review Board (IRB) approval. The keyword "sialendoscopy" was searched on Google in May 2022. The first 100 websites were categorized based on authorship and cataloged in **Figure 1**. The website security level was also documented (https vs http). Websites needed to meet three criteria for inclusion: it was open-access, it was in English text format, and it included sialendoscopy treatment information. Websites containing only videos or slides were excluded since they could not undergo text analysis.

For all included websites, the readability and reliability were measured. Four validated readability metrics were utilized. 5,8,11,12,14 These included the Flesch Reading Ease score (FRES), Flesch-Kincaid Grade Level (FKGL), Gunning Fog (GF) index, and Simple Measure of Gobbledygook (SMOG) index. 15-18 All these metrics consider sentence length, complexity, and vocabulary difficulty to return a score index. A summary of the metrics is included in **Table 1**. 8,11,14-20

Readability scores were captured for the landing page using WebFX's online readability tool. ¹⁴ Since the first 10 seconds spent by the user on a webpage are critical to an individual's likelihood of staying on the website, ²¹ the landing page text was used primarily for measurement. This decision was made because it is unlikely that readers spend more time searching through a website if their first impression of the landing page is not easily readable.

Two validated metrics were used to determine the reliability of website information. ^{6,8,11,22} These included the Discern instrument and the AMA Core Standards for the quality of medical information. Designed at Oxford University, Discern assesses the quality of treatment content for patients and health professionals. ^{6,8} This instrument is a 16-question assessment with overall scores ranging between 16 and 80 points. A higher score indicates better quality content. ⁷ The AMA Core Standards uses four criteria to determine reliability. These include authorship, attribution (references and sources), disclosure (website ownership), and currency (publication date for website). This metric is scored on a 4-point scale. ²²

According to Advanced Web Ranking, individuals using Google's search engine rarely scroll past the first page of search results.²³ The first nine websites returned by Google accounted for 87% of web traffic. Furthermore, roughly 67% of all clicks involved only the first three websites.²³ Therefore, the readability and reliability metrics were compared between the first nine websites (noted as the "top search results") in this study and all websites meeting

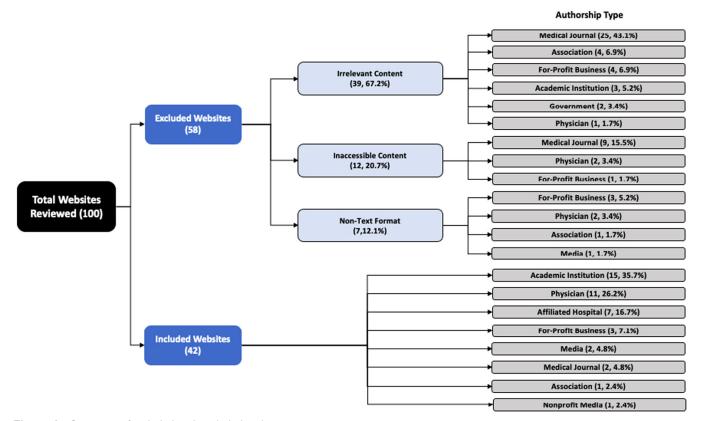


Figure 1. Overview of included and excluded websites.

Singh et al. 3 of 8

Table I. Overview of Readability Tools ^{8,11,14-20}	aadability Tools ^{8,11,14-20}			
Metric	Score approach formula	Score	Score calibration	Score index to other references
Flesch Reading Ease	$206.835 - 1.015 \times (words/sentences) - 84.6 \times (syllables/words)$	0-100	0-30: very difficult, 30-50: difficult, 50-60: fairly difficult, 60-70: standard, 70-80: fairly easy, 80-90: easy, 90-100:	Time magazine = 52, Harvard Law Review = 30
Flesch-Kincaid Grade	$0.39 \times (words/sentences) + 11.8 \times (syllables/words) - 15.59$	0-18+	very easy For the first time reading, it is the grade level needed to understand the text	Romance novels = 6th grade, New York Times = 10th grade
Gunning Fog index	$0.4 \times ([words/sentences] + 100 \times [complexWords/words])$	0-20+	6: sixth grade, 7: seventh grade, 8: eighth grade, 9-12: high school, 13-17: college, 17+: postgraduate	Bible = 6, Time magazine = 10
SMOG index	I.0430 x sqrt (30 x complex)Yords/sentences) + 3.1291	3-18	4.9 and below: elementary school, 5-8.9: middle school, 9-12.9: high school, 13-16.9: undergraduate, 17+: graduate	Article in Reader's Digest = 9, article in Science journal = 15

Abbreviation: SMOG, Simple Measure of Gobbledygook

the inclusion criteria. Differences were assessed for statistical significance using a two sample t test. The purpose of this analysis was to understand if the top search results had superior readability or content quality.

To explore the relationship between readability and reliability, a scatter plot and best-fit line were created for the included websites. FRES was used as the primary metric to represent readability, whereas Discern was chosen as the primary metric to represent quality.

Finally, readability and reliability scores generated were compared with similar studies reviewing other common otolaryngology procedures, such as adenoidectomy, tonsillectomy and sleep apnea, and nasal septoplasty. 8-10 This analysis allowed us to assess trends across the otolaryngology domain and understand if there was more work to be done to improve content overall.

Results

Out of 100 websites reviewed, 42 met the inclusion criteria detailed in **Figure 1**. In addition, 40 of the 42 included websites (95.2%) were considered safe and secure. Among the top search results, six met the inclusion criteria. These websites were classified by authorship: academic institutions (4), physician (1), and association website (1).

Assessing Readability

Multiple readability assessments were performed for the included websites and top search results. The mean FRES for included websites and top search results was 36.2 (standard deviation = 14.2) and 39.5 (standard deviation = 11.5), respectively, with a *P*-value of .543. Both scores aligned with the "difficult to read" category.

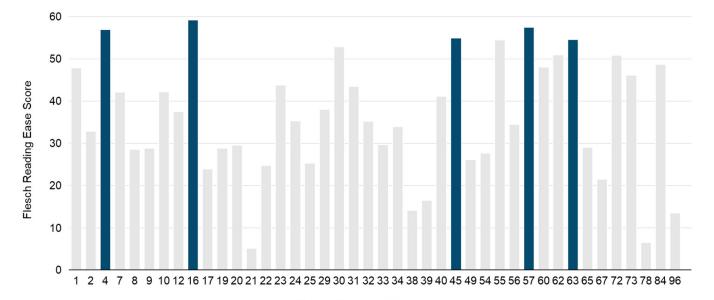
In terms of the grade-level readability metrics, the included websites fell above the suggested sixth-grade level. The mean FKGL for included websites and top search results was 12.9 (standard deviation = 3.0) and 12.0(standard deviation = 3.0), respectively, P-value of .516. The mean GF score for included websites and top search results was 15.9 (standard deviation = 3.4) and 15.2 (standard deviation = 3.8), respectively, with a P-value of .684. Both scores corresponded to an above 12th-grade reading level. Finally, the mean SMOG score for included websites and top search results was 11.9 (standard deviation = 2.3) and 11.2 (standard deviation = 2.3), respectively, with a P-value of .510. Both scores corresponded to a high school reading level (Table 2).

Finally, to assess the ease of readability of the included websites, additional analyses were performed. The five highest-ranking websites were identified using FRES. The scores ranged from 54.5 to 59.2, corresponding to the "fairly difficult" readability category (**Figure 2**). The analysis also revealed that across the five highest FRES websites, only one, the British Association of Oral and Maxillofacial Surgeons, was among the top search results (ranked fourth in the first nine websites).

Table 2. Grade-Level Readability Metric Results

Analysis type	Metric	Mean grade level	Standard deviation	Count of websites below sixth-grade level	% of websites below sixth-grade level
All included websites	Flesch-Kincaid Grade	12.9	3.0	0	0.0
	Gunning Fog index	15.9	3.4	0	0.0
	SMOG index	11.9	2.3	0	0.0
Top search results	Flesch-Kincaid Grade	12.0	3.0	0	0.0
	Gunning Fog index	15.2	3.8	0	0.0
	SMOG index	11.2	2.3	0	0.0

Abbreviation: SMOG, Simple Measure of Gobbledygook.



Rank Order of Included Websites

Figure 2. Flesch Reading Ease Score in search rank order for websites.

Table 3. Discern Score Interpretation

Score		All included websites		Top search results	
classification	Score range	Count	% total	Count	% total
Excellent	63-80	0	0.0	0	0.0
Good	51-63	3	7.1	1	16.7
Fair	39-50	17	40.5	5	83.3
Poor	27-38	18	42.9	0	0.0
Very poor	16-26	4	9.5	0	0.0

Assessing Reliability

Website reliability was assessed based on the landing page text. The mean Discern score for included websites and top search results was 36.9 (standard deviation = 8.9) and 45.0 (standard deviation = 6.7), respectively, with a

P-value of .030. The mean Discern score for included websites aligned with the "poor" category, whereas the top search results' mean Discern score aligned with the "fair" category. Among the included websites, 4 were rated "very poor" (9.5%), 18 fell under the "poor" category (42.9%), 17 were "fair" websites (40%), and 3 were considered "good" (7.1%). In comparison, for the Discern scores of the top search results, 5 (83.3%) of them were "fair" and 1 (16.7%) was classified as "good" (**Table 3**).

The five websites with the highest Discern scores ranged from 48 to 58, corresponding to the "fair" (two websites) and "good" (three websites) categories (**Figure 3**). In addition, across the five highest-scoring websites on Discern, only one, the British Association of Oral and Maxillofacial Surgeons, was among the top search results (ranked fourth in the first nine websites).

The mean AMA benchmark criteria scores for included websites and top search results were 1.8 out of

Singh et al. 5 of 8

4 (standard deviation = 1.1) and 1.5 out of 4 (standard deviation = 0.8), respectively, with a *P*-value of .439. Three included websites (7.1%) scored a 4 out of 4, which is the highest AMA rating possible. Among the three websites with a "good" Discern score, one website scored the highest AMA rating of 4.

Comparing Readability and Reliability

To assess the relationship between the readability (FRES) and reliability (Discern), a best-fit line approach was used. The analysis concluded that there was a very weak positive relationship with a correlation coefficient of 0.063 and an R^2 value of 0.004 between the two variables (**Figure 4**).

Looking at Readability and Reliability Across Similar Otolaryngology Studies

When looking at the adenoidectomy, tonsillectomy and sleep apnea, and nasal septoplasty studies, the mean readability for FRES and FKGL was 48.4 and 10.8, respectively. These scores aligned with the "difficult to read" FRES category and high school reading FKGL. There was no statistical significance found between the mean of the three otolaryngology procedures described above versus the mean sialendoscopy score for all included websites (*P*-value of .15 for FRES comparison, *P*-value of .13 for FKGL comparison).

In terms of reliability, the mean Discern score across the three otolaryngology procedures was 47.0,

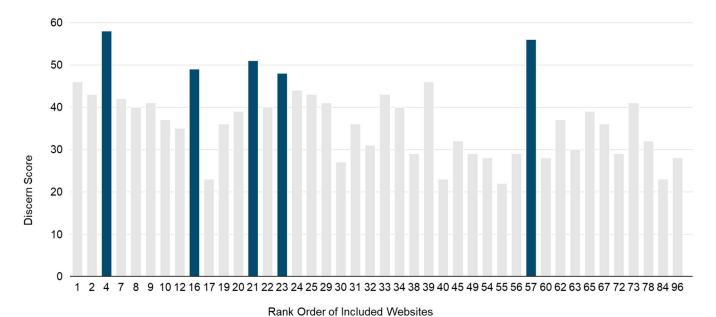


Figure 3. Discern instrument score in search rank order for websites.

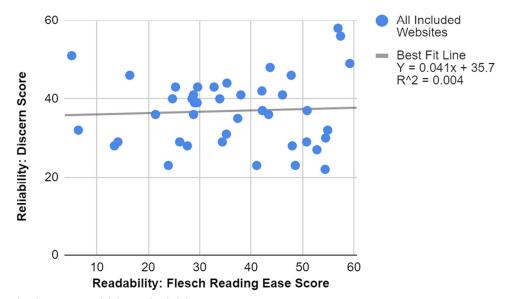


Figure 4. Relationship between readability and reliability metrics.

corresponding to the "fair" category. Similarly, there was no statistical significance found between the mean of the three otolaryngology procedures versus the mean sialendoscopy score for all included websites (*P*-value of .06 for Discern comparison) (**Table 4**).

Discussion

This study's findings suggest that otolaryngologists can improve both the readability and reliability of online health information on sialendoscopy. The low mean readability ("difficult to read" category) and reliability ("poor" category) scores highlight that online health information on sialendoscopy is not easily understandable for the patient population. The most readable and highest quality websites were not the highest ranked on Google's search engine. Furthermore, the weak positive correlation coefficient between readability and reliability demonstrates the limited relationship between the two variables and illustrates that search engine algorithms could be improved for health information seekers.

Given that the first page of search results accounts for the vast majority of web traffic, ²³ the study also sought to understand whether the top search results websites had statistically superior readability or content quality compared to all included websites. Although there was no statistically significant difference in readability scores between top search results and included websites, the results showed a statistically significant difference for the primary reliability metric (Discern). The mean score for the top search results was 8.1 points higher than the mean score for all included websites. The reliability scores may be statistically better for the top search results because Google's algorithm prioritizes higher-quality scoring websites in its search results. Google's algorithm looks at five factors to determine which results rise to the top of the query: meaning, relevance, quality, usability, and context. Specifically, Google focuses on "content that demonstrates expertise, authoritativeness, and trustworthiness."²⁴ In comparison, readability is not a key factor like reliability on Google.

Unfortunately, the results of this study align with deficiencies found in similar studies reviewing readability and reliability of other common otolaryngology procedures, including adenoidectomy, tonsillectomy and sleep apnea, and nasal septoplasty. 8-10 The readability scores across all of these studies not only corresponded to "difficult" and "fairly difficult" reading levels, but the mean FKGL was also well above the suggested sixth-grade reading level. Although there was a categorical difference between the mean Discern score of the three otolaryngology procedures and sialendoscopy ("fair" vs "poor"), there was no statistical difference. This categorical difference between the Discern scores could be due to the subjective nature of scoring for the 16-point metric.

One reason for the "difficult" and "fairly difficult" readability scores across otolaryngology procedures could be because medical terminology is inherently complex. Words such as "sialendoscopy," "procedure," "glands," "inflammation," and "blockages" may be difficult to comprehend for health information seekers. Research has found that the readability of health information is significantly easier when medical terminology is removed.²⁵ There is an opportunity for otolaryngologists to collaborate with medical writers and communication specialists to use simpler and less complex words without impacting quality. Authors should proactively consider using readability and reliability tools to test their content before publication. This type of deliberate effort would ensure that the health information is comprehensible for otolaryngology patients.

With the advent of artificial intelligence (AI), chatbots could have a profound impact on future medical research and publications. Large language models are able to write essays, draft papers, simplify text to specific reading levels, and summarize literature concisely. In addition, AI could help accelerate the innovation process by shortening time-to-publication and the fluidity of the writer's language.²⁶ Although this research did not look into AI's impact on online medical education, it could be valuable to explore in future studies. However, before its use, three challenges must be addressed to create the most optimal results. One, sources must be identified from the chatbot's response since chatbots can be fundamentally unreliable at answering questions, occasionally generating false responses. 26,27 Second, the prompt used to seek patient education must be phrased to generate the most readable and reliable information about the queried topic. Lastly, the chatbot must provide a response below a sixth-grade reading level without sacrificing the quality of the

Table 4. Comparison of Readability and Reliability Across Similar Otolaryngology Procedures⁸⁻¹⁰

		Tonsillectomy and sleep		Mean across otolaryngology	
Metric	Adenoidectomy	apnea	Nasal septoplasty	procedures	Sialendoscopy
Flesch-Kincaid Reading Ease	48.1 (difficult)	42.3 (difficult)	54.9 (fairly difficult)	48.4 (difficult)	36.2 (difficult)
Flesch-Kincaid Grade Level	10.8	10.7	9.2ª	10.8	12.9
Discern	43.0 (fair)	55.1 (good)	42.9 (fair)	47.0 (fair)	36.9 (poor)

 $^{^{\}mathrm{a}}\mathrm{Data}$ estimated from **Figure 1** bar chart in the nasal septoplasty study. $^{\mathrm{10}}$

Singh et al. 7 of 8

information. Important content aspects (eg, description of risks, impact on quality of life, and alternative options) should not be weakened at the cost of readability.

Five limitations were identified in this study. Our analysis focused on websites returned only by Google's search engine based on our search history, personal data, and interests. This decision could result in two possible limitations. First, although Google has the highest search engine market share worldwide across all platforms (92.4% as of October 2022), there are other search engine options available for online health seekers. Second, Google has stated that it personalized searches based on user activity and frequently updates its search engine algorithm. Therefore, the websites analyzed for the "top search results" may differ from those generated for other health seekers.

In addition, there are other ways of spelling the word "sialendoscopy." For example, "sialoendoscopy" is also an alternate spelling for this procedure. However, for the purposes of this study, only the first spelling was used in our search. Also, although 100 websites were retrieved for the initial analysis, less than half of the websites met the inclusion criteria, given that they were either behind a paywall, displayed in a nontext format, or contained content unrelated to sialendoscopy treatment. This limited our review of the readability and reliability of websites related to sialendoscopy. Lastly, when assessing reliability metrics, scoring was based on a subjective assessment by the researcher. Unlike the readability metrics, purely formulaic systems were not used. Therefore, reliability measurements were solely based on the interpretation and judgment of the authors, which may have impacted our results.

Conclusions

The low readability and reliability scores implied that the online health information on sialendoscopy is not easily understandable at a reading level appropriate for the general public.

Our findings showed that the most readable and highest quality websites were not the highest ranked in our search results. Because online health information can sway an individual's perception, improving the readability and quality can have a profound effect. Search engine algorithms, which do not prioritize readable content, need to be enhanced for the benefit of health information seekers. Additionally, otolaryngologists should consider simplifying their content and leveraging new technologies, such as AI chatbots, to improve online health information for patients. As AI evolves, future studies should be conducted to assess its impact on the readability and reliability of online health information.

Acknowledgments

Thank you to Aman Singh for his support on the manuscript bibliography, as well as figure and table revisions.

Author Contributions

Rohan Singh, concept, design, acquisition, analysis, interpretation of data, drafting of the manuscript, critical revision of the manuscript for important intellectual content, accountability for all aspects of the work; Christopher Badger, concept, design, acquisition, analysis, interpretation of data, drafting of the manuscript, critical revision of the manuscript for important intellectual content, accountability for all aspects of the work; Arjun S. Joshi, concept, design, critical revision of the manuscript for important intellectual content, accountability for all aspects of the work.

Disclosures

Competing interests: None.

Funding source: None.

ORCID iD

References

- Turner E, Rainie L. Most Americans rely on their own research to make big decisions, and that often means online searches. Pew Research Center. August 21, 2022. https://www.pewresearch.org/fact-tank/2020/03/05/mostamericans-rely-on-their-own-research-to-make-bigdecisions-and-that-often-means-online-searches/
- Author N. The internet and health. Pew Research Center: Internet, Science & Tech. February 12, 2013. https://www.pewresearch.org/internet/2013/02/12/the-internet-and-health/
- Hammett JT, Walker C. Sialolithiasis. In: StatPearls. StatPearls Publishing; 2022. http://www.ncbi.nlm.nih.gov/books/NBK549845/
- Erkul E, Cekin E, Gungor A. Long-term outcomes of sialendoscopy in the management of sialolithiasis and idiopathic chronic sialadenitis with ductal scars. *Turk Arch Otorhinolaryngol*. 2019;57(2):75-80. doi:10.5152/tao.2019.4290
- Kher A, Johnson S, Griffith R. Readability assessment of online patient education material on congestive heart failure. *Adv Prev Med.* 2017;2017:9780317. doi:10.1155/2017/9780317
- 6. Fahy E. Quality of patient health information on the internet: reviewing a complex and. *Australas Med J.* 2014;7(1):24-28. doi:10.4066/AMJ.2014.1900
- 7. Discern—welcome to discern. August 21, 2022. http://www.discern.org.uk/index.php
- 8. Ahmadi O, Wood AJ. The readability and reliability of online information about adenoidectomy. *J Laryngol Otol*. 2021;135(11):976-980. doi:10.1017/S0022215121002164
- Chi E, Jabbour N, Aaronson NL. Quality and readability of websites for patient information on tonsillectomy and sleep apnea. *Int J Pediatr Otorhinolaryngol*. 2017;98:1-3. doi:10. 1016/j.ijporl.2017.04.031
- Grose EM, Holmes CP, Aravinthan KA, Wu V, Lee JM. Readability and quality assessment of internet-based patient education materials related to nasal septoplasty. J

Otolaryngol Head Neck Surg. 2021;50(1):16. doi:10.1186/s40463-021-00507-z

- 11. Grewal P, Alagaratnam S. The quality and readability of colorectal cancer information on the internet. *Int J Surg.* 2013;11(5):410-413. doi:10.1016/j.ijsu.2013.03.006
- 12. Patel CR, Cherla DV, Sanghvi S, Baredes S, Eloy JA. Readability assessment of online thyroid surgery patient education materials. *Head Neck*. 2013;35(10):1421-1425. doi:10.1002/hed.23157
- 13. Encrypting the web. Electronic Frontier Foundation. December 7, 2022. https://www.eff.org/encrypt-the-web
- 14. Readability test. webfx. August 21, 2022. https://www.webfx.com/tools/read-able/
- 15. Flesch R. How to Write Plain English: A Book for Lawyers and Consumers. 1st ed. Harper & Row; 1979.
- Flesch R. A new readability yardstick. J Appl Psychol. 1948;32(3):221-233. doi:10.1037/h0057532
- 17. Gunning R. *The Technique of Clear Writing*. Mcgraw-Hill; 1952.
- 18. McLaughlin GH. SMOG grading: a new readability formula. *J Read*. 1969;22:639-664.
- 19. Smog readability calculator—Smog index. August 21, 2022. https://charactercalculator.com/smog-readability/
- How do you measure readability? full media. August 21, 2022. https://www.fullmedia.com/how-do-you-measure-readability
- 21. Experience, W. L. in R.-B. U. How long do users stay on web pages? Nielsen Norman Group. November 13, 2022.

- https://www.nngroup.com/articles/how-long-do-users-stay-on-web-pages/
- Silberg WM. Assessing, controlling, and assuring the quality of medical information on the internet: Caveant lector et viewor let the reader and viewer beware. *JAMA*. 1997;277(15): 1244-1245. doi:10.1001/jama.1997.03540390074039
- 23. Google organic CTR history—advanced web ranking. August 21, 2022. https://www.advancedwebranking.com/ctrstudy/
- 24. Ranking results—how Google search works. Google Search—Discover How Google Search Works. March 26, 2023. https://www.google.com/search/howsearchworks/how-search-works/ranking-results/
- 25. Sand-Jecklin K. The impact of medical terminology on readability of patient education materials. *J Community Health Nurs*. 2007;24(2):119-129. doi:10.1080/07370010701316254
- van Dis EAM, Bollen J, Zuidema W, van Rooij R, Bockting CL. ChatGPT: five priorities for research. *Nature*. 2023;614(7947):224-226. doi:10.1038/d41586-023-00288-7
- 27. Stokel-Walker C, Van Noorden R. What ChatGPT and generative AI mean for science. *Nature*. 2023;614(7947): 214-216. doi:10.1038/d41586-023-00340-6
- 28. Personalization & Google search results—Google search help. March 7, 2025. https://support.google.com/websearch/answer/12410098?hl=en
- 29. Search engine market share worldwide. StatCounter Global Stats. November 13, 2022. https://gs.statcounter.com/search-engine-market-share