Flesch-Kincaid Readability Scores and Audience Comprehension of Complex Health Information

1. Introduction

The Flesch-Kincaid readability formulas are widely used to estimate the reading grade level required to understand health information, with the goal of ensuring that patient education materials are accessible to the general public. However, a substantial body of research demonstrates that most health information—whether online, in pamphlets, or in consent forms—is written at a reading level significantly higher than recommended, often exceeding the average literacy of the intended audience (Walsh & Volsko, 2008; Degen et al., 2025; Boutemen & Miller, 2023; Kue et al., 2021; Huang et al., 2015; Roberts et al., 2016; Williams et al., 2016; Abdullah et al., 2022; Armache et al., 2024; Singh et al., 2025; Crabtree & Lee, 2022). Studies consistently find that materials with higher Flesch-Kincaid grade levels are less likely to be comprehended by patients, especially those with limited health literacy, and that reducing the grade level improves accessibility and understanding (Jindal & Macdermid, 2017; Ley & Florio, 1996; Williams et al., 2016; Espinosa et al., 2022). Despite these findings, the majority of health information materials remain too complex, highlighting a persistent gap between readability scores and actual patient comprehension (Walsh & Volsko, 2008; Boutemen & Miller, 2023; Kue et al., 2021; Huang et al., 2015; Roberts et al., 2016; Williams et al., 2016; Abdullah et al., 2022; Armache et al., 2024; Singh et al., 2025; Crabtree & Lee, 2022). Furthermore, while Flesch-Kincaid scores provide a useful benchmark, they do not account for all factors influencing comprehension, such as prior knowledge, motivation, or the presence of visual aids (Jindal & Macdermid, 2017; Wang et al., 2013; Aleligay et al., 2008; Williams et al., 2016). This review synthesizes the evidence on the relationship between Flesch-Kincaid readability scores and audience comprehension of complex health information, examining both the strengths and limitations of these metrics.

2. Methods

A comprehensive literature search was conducted across over 170 million research papers in Consensus, including sources such as Semantic Scholar and PubMed. The search strategy targeted studies evaluating the relationship between Flesch-Kincaid readability scores and comprehension of health information, as well as critiques and comparative analyses of readability formulas. A total of 1,034 papers were identified, 590 were screened after de-duplication, 468 met eligibility criteria, and the top 50 most relevant papers were included in this review.

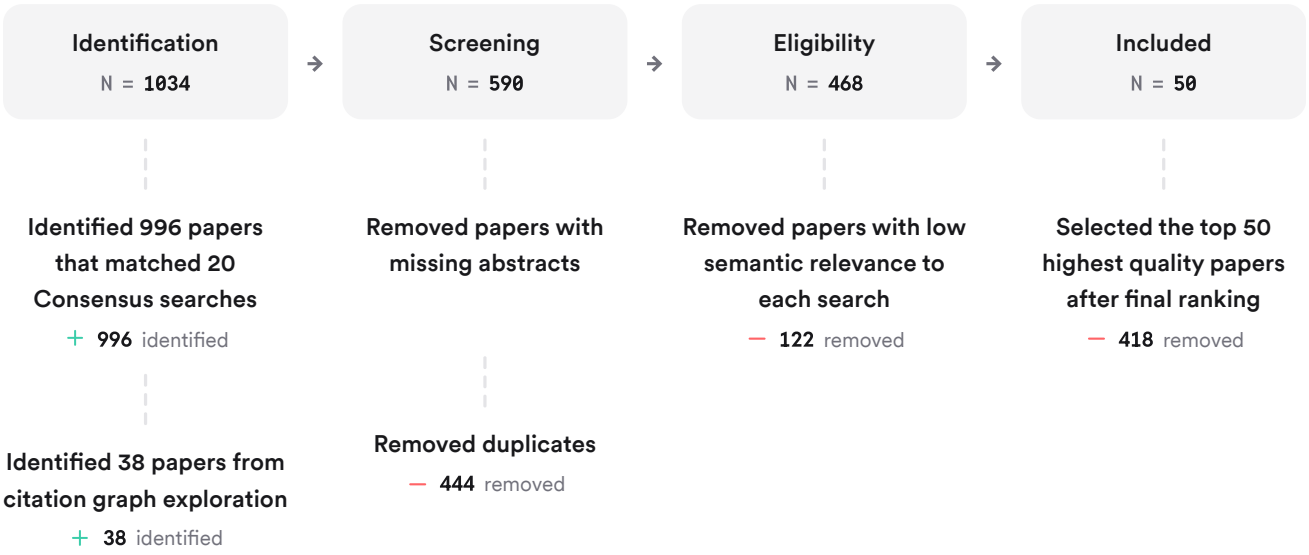

FIGURE 1 Flow diagram of search and selection process.

Twenty unique search strategies were used, focusing on readability, comprehension, and health literacy in patient education materials.

3. Results

3.1 Readability Levels of Health Information

Multiple studies report that the average Flesch-Kincaid grade level of health information materials is well above the recommended sixth- to eighth-grade level, often ranging from 9th to 14th grade or higher (Walsh & Volsko, 2008; Degen et al., 2025; Boutemen & Miller, 2023; Kue et al., 2021; Huang et al., 2015; Roberts et al., 2016; Williams et al., 2016; Abdullah et al., 2022; Armache et al., 2024; Singh et al., 2025; Crabtree & Lee, 2022). For example, online materials for conditions such as stroke, diabetes, cancer, and mental health frequently require high school or college-level reading skills (Degen et al., 2025; Walsh & Volsko, 2008; Boutemen & Miller, 2023; Kue et al., 2021; Huang et al., 2015; Roberts et al., 2016; Williams et al., 2016; Abdullah et al., 2022; Armache et al., 2024; Singh et al., 2025; Crabtree & Lee, 2022).

3.2 Impact on Audience Comprehension

There is a strong inverse relationship between Flesch-Kincaid grade level and patient comprehension: as the grade level increases, comprehension decreases, particularly among individuals with limited health literacy (Jindal & Macdermid, 2017; Ley & Florio, 1996; Williams et al., 2016; Espinosa et al., 2022; Armache et al., 2024). Studies show that materials rewritten to lower grade levels (closer to sixth grade) are more likely to be understood by a broader audience (Williams et al., 2016; Espinosa et al., 2022; Ellison et al., 2025). However, even when materials meet recommended grade levels, other factors such as medical jargon, sentence complexity, and lack of visual aids can still hinder comprehension (Jindal & Macdermid, 2017; Wang et al., 2013; Aleligay et al., 2008; Williams et al., 2016).

3.3 Limitations of Flesch-Kincaid and Other Readability Formulas

While Flesch-Kincaid scores are useful for benchmarking, they do not account for document layout, visual aids, prior knowledge, or cultural context, all of which influence comprehension (Jindal & Macdermid, 2017; Wang et al., 2013; Aleligay et al., 2008; Williams et al., 2016). Some studies find that Flesch-Kincaid may underestimate the true reading difficulty compared to other formulas like SMOG, which is often considered more stringent for health materials (Grabeel et al., 2018; Wang et al., 2013; Sharma et al., 2014; Wilson, 2009).

3.4 Interventions and Improvements

Interventions such as using plain language, reducing sentence length, and leveraging AI tools (e.g., ChatGPT) to rewrite materials at lower grade levels have been shown to significantly improve readability and, by extension, potential comprehension (Kianian et al., 2023; Oliva et al., 2024; Ellison et al., 2025). However, the effectiveness of these interventions depends on careful implementation and ongoing evaluation (Kianian et al., 2023; Oliva et al., 2024; Ellison et al., 2025).

Key Papers

Paper	Methodology	Sample Size	Key Results
(Walsh & Volsko, 2008)	Cross-sectional analysis of 100 online articles	100 articles	Most materials exceeded 7th-grade level; higher grade level linked to lower comprehension
(Jindal & Macdermid, 2017)	Critical review	N/A	Flesch-Kincaid does not account for all comprehension factors; useful but limited
(Williams et al., 2016)	Systematic review and institutional intervention	950 materials (13 studies) + 12 handouts	Revising materials from 10th to 6th grade improved readability and suitability
(Wang et al., 2013)	Comparative analysis of readability formulas	15 samples	Flesch-Kincaid most used; SMOG more stringent; up to 5-grade variability in scores
(Ellison et al., 2025)	Experimental comparison of AI-generated vs. baseline materials	52 topics	AI-generated materials with targeted prompts achieved recommended grade levels

FIGURE 2 Comparison of key studies on Flesch-Kincaid readability and comprehension of health information.

Top Contributors

Type	Name	Papers
Author	Andrew M. Williams	(Williams et al., 2016)
Author	Tiffany M Walsh	(Walsh & Volsko, 2008)
Author	Lih-Wern Wang	(Wang et al., 2013)
Journal	*JAMA ophthalmology*	(Huang et al., 2015)
Journal	*The Journal of bone and joint surgery. American volume*	(Roberts et al., 2016)
Journal	*BMC Ophthalmology*	(Williams et al., 2016)

FIGURE 3 Authors & journals that appeared most frequently in the included papers.

4. Discussion

The evidence overwhelmingly indicates that most health information is written at a level too advanced for the average patient, which can impede comprehension and informed decision-making (Walsh & Volsko, 2008; Degen et al., 2025; Boutemen & Miller, 2023; Kue et al., 2021; Huang et al., 2015; Roberts et al., 2016; Williams et al., 2016; Abdullah et al., 2022; Armache et al., 2024; Singh et al., 2025; Crabtree & Lee, 2022). Lowering the Flesch-Kincaid grade level of materials is associated with improved accessibility and potential comprehension, but this alone is not sufficient. Readability formulas like Flesch-Kincaid provide a useful, standardized metric, but they do not capture all the nuances of understanding, such as the impact of visual aids, prior knowledge, or cultural context (Jindal & Macdermid, 2017; Wang et al., 2013; Aleligay et al., 2008; Williams et al., 2016). Furthermore, Flesch-Kincaid often underestimates reading difficulty compared to more stringent formulas like SMOG, suggesting that even materials meeting Flesch-Kincaid recommendations may still be too complex for some audiences (Grabeel et al., 2018; Wang et al., 2013; Sharma et al., 2014; Wilson, 2009).

Recent advances, such as the use of AI tools to generate or revise patient education materials, show promise in improving readability, but require careful prompt engineering and validation to ensure accuracy and appropriateness (Kianian et al., 2023; Oliva et al., 2024; Ellison et al., 2025). Ultimately, improving comprehension of complex health information requires a multifaceted approach that goes beyond readability scores, incorporating plain language, visual aids, and user testing with target populations (Jindal & Macdermid, 2017; Williams et al., 2016; Ellison et al., 2025).

Claims and Evidence Table

Claim	Evidence Strength	Reasoning	Papers
Most health information materials exceed recommended Flesch-Kincaid grade levels	Strong	Consistent findings across multiple large-scale studies and systematic reviews	(Walsh & Volsko, 2008; Degen et al., 2025; Boutemen & Miller, 2023; Kue et al., 2021; Huang et al., 2015; Roberts et al., 2016; Williams et al., 2016; Abdullah et al., 2022; Armache et al., 2024; Singh et al., 2025; Crabtree & Lee, 2022)
Lowering Flesch-Kincaid grade level improves accessibility and potential comprehension	Strong	Interventional studies show improved readability and suitability after revision	(Williams et al., 2016; Espinosa et al., 2022; Ellison et al., 2025)
Flesch-Kincaid does not account for all factors influencing comprehension	Moderate	Critical reviews highlight missing elements (visuals, prior knowledge, etc.)	(Jindal & Macdermid, 2017; Wang et al., 2013; Aleligay et al., 2008; Williams et al., 2016)
Flesch-Kincaid often underestimates reading difficulty compared to SMOG	Moderate	Comparative studies show SMOG yields higher grade levels	(Grabeel et al., 2018; Wang et al., 2013; Sharma et al., 2014; Wilson, 2009)
AI tools can improve readability when properly prompted	Moderate	Early studies show AI-generated materials can meet grade-level targets	(Kianian et al., 2023; Oliva et al., 2024; Ellison et al., 2025)
Readability alone does not guarantee comprehension or informed decision-making	Moderate	Comprehension depends on multiple factors beyond text complexity	(Jindal & Macdermid, 2017; Williams et al., 2016; Espinosa et al., 2022; Armache et al., 2024)

FIGURE Key claims and support evidence identified in these papers.

5. Conclusion

The relationship between Flesch-Kincaid readability scores and audience comprehension of complex health information is clear: lower grade levels are associated with improved accessibility, but true comprehension depends on a broader set of factors. Most health information materials remain too complex for the average reader, underscoring the need for ongoing efforts to simplify language, incorporate user feedback, and use multiple strategies to enhance understanding.

5.1 Research Gaps

Despite extensive research on readability, there are gaps in understanding how best to translate improved readability into real-world comprehension and health outcomes, especially for diverse populations and in digital contexts.

Research Gaps Matrix

Topic/Outcome	General Population	Low Health Literacy	Digital Materials	Visual Aids Included	AI-Generated Materials
Readability Assessment	38	12	22	7	5
Comprehension Testing	15	6	8	3	2
Intervention Studies	10	4	5	2	3
Longitudinal Outcomes	2	1	1	GAP	GAP

FIGURE Matrix of research topics and study attributes, highlighting areas with limited research.

5.2 Open Research Questions

Future research should focus on how improved readability translates to actual comprehension and health outcomes, especially in diverse and low-literacy populations, and how digital and AI-generated materials can be optimized for understanding.

Question	Why
How does lowering Flesch-Kincaid grade level affect real-world comprehension and health outcomes in diverse populations?	Understanding this link is crucial for designing effective patient education that leads to better health behaviors and outcomes.
What additional factors (e.g., visuals, cultural tailoring) most enhance comprehension beyond readability scores?	Identifying these factors can help create more effective, inclusive health information for all literacy levels.
How can AI tools be optimized to generate both readable and accurate health information for patients?	Ensuring AI-generated materials are both accessible and reliable is vital as these tools become more widely used in healthcare.

FIGURE Open research questions for future investigation on readability and comprehension.

In summary, while Flesch-Kincaid readability scores are a valuable tool for assessing the accessibility of health information, true comprehension requires a multifaceted approach that addresses the complexity of both language and audience needs.

These papers were sourced and synthesized using Consensus, an AI-powered search engine for research. Try it at https://consensus.app

References

Kianian, R., Sun, D., Crowell, E., & Tsui, E. (2023). The Use of Large Language Models to Generate Education Materials About Uveitis.. *Ophthalmology. Retina*. https://doi.org/10.1016/j.oret.2023.09.008

Degen, N., Sivakumar, M., Varkey, T., Alexandrov, A., & Singh, S. (2025). Abstract DP20: An Evaluation of Stroke Literature Pamphlets For Stroke Patients. *Stroke*. https://doi.org/10.1161/str.56.suppl_1.dp20

Jindal, P., & Macdermid, J. (2017). Assessing reading levels of health information: uses and limitations of flesch formula. *Education for Health*, 30, 84 - 88. https://doi.org/10.4103/1357-6283.210517

Walsh, T., & Volsko, T. (2008). Readability assessment of internet-based consumer health information.. *Respiratory care*, 53 10, 1310-5.

Grabeel, K., Russomanno, J., Oelschlegel, S., Tester, E., & Heidel, R. (2018). Computerized versus hand-scored health literacy tools: a comparison of Simple Measure of Gobbledygook (SMOG) and Flesch-Kincaid in printed patient education materials. *Journal of the Medical Library Association : JMLA*, 106, 38 - 45. https://doi.org/10.5195/jmla.2018.262

Boutemen, L., & Miller, A. (2023). Readability of publicly available mental health information: A systematic review.. *Patient education and counseling*, 111, 107682. https://doi.org/10.1016/j.pec.2023.107682

Kue, J., Klemanski, D., & Browning, K. (2021). Evaluating Readability Scores of Treatment Summaries and Cancer Survivorship Care Plans. *JCO Oncology Practice*, 17, 615 - 621. https://doi.org/10.1200/OP.20.00789

Wang, L., Miller, M., Schmitt, M., & Wen, F. (2013). Assessing readability formula differences with written health information materials: application, results, and recommendations.. *Research in social & administrative pharmacy : RSAP*, 9 5, 503-16. https://doi.org/10.1016/j.sapharm.2012.05.009

Aleligay, A., Worrall, L., & Rose, T. (2008). Readability of written health information provided to people with aphasia. *Aphasiology*, 22, 383 - 407. https://doi.org/10.1080/02687030701415872

Huang, G., Fang, C., Agarwal, N., Bhagat, N., Eloy, J., & Langer, P. (2015). Assessment of online patient education materials from major ophthalmologic associations.. *JAMA ophthalmology*, 133 4, 449-54. https://doi.org/10.1001/jamaophthalmol.2014.6104

Ley, P., & Florio, T. (1996). The use of readability formulas in health care. *Psychology Health & Medicine*, 1, 7-28. https://doi.org/10.1080/13548509608400003

Roberts, H., Zhang, D., & Dyer, G. (2016). The Readability of AAOS Patient Education Materials: Evaluating the Progress Since 2008.. *The Journal of bone and joint surgery. American volume*, 98 17, e70. https://doi.org/10.2106/JBJS.15.00658

Williams, A., Muir, K., & Rosdahl, J. (2016). Readability of patient education materials in ophthalmology: a single-institution study and systematic review. *BMC Ophthalmology*, 16. https://doi.org/10.1186/s12886-016-0315-0

Oliva, A., Pasick, L., Hoffer, M., & Rosow, D. (2024). Improving readability and comprehension levels of otolaryngology patient education materials using ChatGPT.. *American journal of otolaryngology*, 45 6, 104502. https://doi.org/10.1016/j.amjoto.2024.104502

Sharma, N., Tridimas, A., & Fitzsimmons, P. (2014). A readability assessment of online stroke information.. *Journal of stroke and cerebrovascular diseases : the official journal of National Stroke Association*, 23 6, 1362-7. https://doi.org/10.1016/j.jstrokecerebrovasdis.2013.11.017

Espinosa, J., Lucerna, A., & Schuitema, H. (2022). Healthcare Documents: How Readable? A Study of Simple Strategies to Decrease Reading Grade Level and Increase Reading Ease of Healthcare Documents. *Mathews Journal of Emergency Medicine*. https://doi.org/10.30654/mjem.10044

Abdullah, Y., Alokozai, A., O'Connell, S., & Mulcahey, M. (2022). Online Patient Education Materials for Common Sports Injuries Are Written at Too-High of a Reading Level: A Systematic Review. *Arthroscopy, Sports Medicine, and Rehabilitation*, 4, e861 - e875. https://doi.org/10.1016/j.asmr.2021.12.017

Armache, M., Assi, S., Wu, R., Jain, A., Lu, J., Gordon, L., Jacobs, L., Fundakowski, C., Rising, K., Leader, A., Fakhry, C., & Mady, L. (2024). Readability of Patient Education Materials in Head and Neck Cancer: A Systematic Review.. *JAMA otolaryngology-- head & neck surgery*. https://doi.org/10.1001/jamaoto.2024.1569

Singh, A., Gupta, N., Chien, D., Singh, R., Sachdeva, A., Danasek, K., Gajjar, A., Houk, C., Jalal, M., Li, A., Whyte, R., Stone, J., & Vates, G. (2025). 1363 Assessing the Readability of Patient Education Materials for Common Neurosurgical Conditions. *Neurosurgery*, 71, 229 - 229. https://doi.org/10.1227/neu.0000000000003360_1363

Wilson, M. (2009). Readability and Patient Education Materials Used for Low-Income Populations. *Clinical Nurse Specialist*, 23, 33-40. https://doi.org/10.1097/01.NUR.0000343079.50214.31

Ellison, I., Oslock, W., Abdullah, A., Wood, L., Thirumalai, M., English, N., Jones, B., Hollis, R., Rubyan, M., & Chu, D. (2025). De novo generation of colorectal patient educational materials using large language models: Prompt engineering key to improved readability.. *Surgery*, 180, 109024. https://doi.org/10.1016/j.surg.2024.109024

Crabtree, L., & Lee, E. (2022). Assessment of the readability and quality of online patient education materials for the medical treatment of open-angle glaucoma. *BMJ Open Ophthalmology*, 7. https://doi.org/10.1136/bmjophth-2021-000966