NORTHWESTERN MEDICAL JOURNAL

Northwestern Med J. 2026;6(4):308-316
DOI: 10.54307/NWMJ.2026.256

RESEARCH ARTICLE

Readability assessment of ChatGPT-4 and ChatGPT-5 responses to ear care questions

Cemal Özyılmaz[1], **Mehmet Fatih Karakuş**[1]

[1]Department of Otorhinolaryngology, Faculty of Medicine, Ordu University, Ordu, Türkiye

Cite as: Özyılmaz C, Karakuş MF. Readability assessment of ChatGPT-4 and ChatGPT-5 responses to ear care questions. Northwestern Med J. 2026;6(4):308-316.

ABSTRACT

Aim: To assess the readability of ChatGPT-4o and ChatGPT-5 responses to questions about ear care.

Materials and Methods: Ten questions about ear care from the World Health Organization's 'Questions and Answers' section were each posed seven times to both ChatGPT-4o and ChatGPT-5. A total of 70 responses were obtained and analyzed for readability.

Results: For ChatGPT-4o, the mean readability scores were: Flesch Reading Ease 62.84 ± 8.32, Flesch–Kincaid Grade Level 8.13 ± 1.59, Simple Measure of Gobbledygook 7.35 ± 1.42, Gunning Fog Index 8.44 ± 1.79, Coleman–Liau Index 9.62 ± 1.71, and Automated Readability Index 8.70 ± 2.02. For ChatGPT-5, the corresponding scores were: Flesch Reading Ease 72.93 ± 7.69, Flesch–Kincaid Grade Level 5.75 ± 1.17, Simple Measure of Gobbledygook 6.35 ± 0.79, Gunning Fog Index 7.37 ± 1.22, Coleman–Liau Index 7.32 ± 1.52, and Automated Readability Index 5.78 ± 1.42. ChatGPT-5 demonstrated better readability than ChatGPT-4o across all readability indices ($p<0.001$).

Conclusion: The readability of the answers produced by ChatGPT-4o is "standard", and that of ChatGPT-5 is "fairly easy". ChatGPT-5 responses showed improved readability over ChatGPT-4o, aligning more closely with patient education readability standards and underscoring its potential as an accessible source of health information.

Keywords: artificial intelligence, ChatGPT health information, ear care, readability

Corresponding author: Mehmet Fatih Karakuş **E-mail:** mfkarakus@yahoo.com
Received: 09.05.2026 **Accepted:** 24.08.2026 **Published:** 06.10.2026

Copyright © 2026 The Author(s). This is an open-access article published by Bolu Izzet Baysal Training and Research Hospital under the terms of the **Creative Commons Attribution License (CC BY)** which permits unrestricted use, distribution, and reproduction in any medium or format, provided the original work is properly cited.

INTRODUCTION

Ear care is an important part of personal hygiene and plays a fundamental role in maintaining ear function. Behaviours such as inserting foreign objects into the ear, prolonged exposure to loud noise, and swimming in unclean water can adversely affect ear health. If proper ear care is not carried out, various infections and related complications may arise. However, public knowledge and awareness of ear care are insufficient. Impacted cerumen and the associated hearing loss that develop as a result of incorrect ear care practices have become a health issue that requires attention (1-3).

Cerumen, formed by secretions from the ceruminous and sebaceous glands in the external auditory canal, plays an important role in maintaining ear health. A certain amount of cerumen moistens the ear canal, forms a physical barrier against foreign objects, and has antimicrobial properties. However, attempts by individuals to clean their ears with cotton buds or unsuitable tools can disrupt this natural mechanism, leading to trauma to the external auditory canal, damage to the tympanic membrane, cerumen impaction, and infection. Societal awareness of the potential harms associated with the use of cotton buds is generally low. Studies show that individuals have insufficient knowledge about ear cleaning habits and the complications associated with these habits (3,4).

Currently, individuals access health information from various sources, including healthcare professionals, printed materials, online search engines, social media, and health portals. In recent years, with the proliferation of the internet, online search engines, social media, and health portals have become among the most frequently used sources of information for the public (5). However, rapid advances in artificial intelligence have ushered in a new era in the ways individuals access healthcare information. Chat Generative Pre-trained Transformer (ChatGPT) is an artificial intelligence (AI) system developed by OpenAI and based on large language models. It can understand and generate coherent and accurate responses to text-based queries in natural language. By learning the structures of language, ChatGPT can provide information on a wide range of topics and engage in human-like conversations. It is trained on a vast corpus of text data using deep learning techniques to acquire general knowledge. The ability of AI-powered language models such as ChatGPT to provide users with quick, comprehensive and accessible answers on health matters increases the potential for these platforms to become one of the primary sources for seeking health-related information (6,7).

ChatGPT has been shown to perform well in medical examinations and provide accurate information across various medical specialities (8-10). The literature reports that ChatGPT achieves high success rates in answering questions in the ear, nose, and throat (ENT) field. Its success in the audiology competency exam suggests it could be a potential tool in hearing care services (11); it has been reported to provide highly accurate answers to ENT questions from the public (12) and that its diagnostic and treatment approaches in easy to moderately difficult ENT scenarios largely align with those of physicians (13).

Artificial intelligence-based platforms can not only facilitate access to information but also influence individuals' health-related decision-making processes. Making informed health decisions depends not only on the accuracy of the information provided but also on the target audience's ability to understand this information (14). At this point, the concept of ‘readability’ comes to the fore. Readability is defined as the characteristics that facilitate or hindercomprehension of reading material. Readability is generally expressed as the minimum number of years of formal education required to read and understand a text (15). Today, individuals of all educational backgrounds can make health decisions based on information obtained from AI-based platforms such as ChatGPT. Therefore, in addition to the accuracy of ChatGPT's responses, their readability is also critical parameter. This study aimed to evaluate the readability of responses generated by ChatGPT-4o and ChatGPT-5 to questions about ear care, to compare readability between the two models, and to assess their readability relative to the World Health Organization’s (WHO) materials on the same topic. The findings of this study may provide insights into whether newer AI models offer more patient-friendly health information, thereby supporting the informed use of large language models in healthcare settings.

MATERIALS AND METHODS

This study does not involve any research conducted with human or animal subjects, and ChatGPT, used in this study, is a publicly available application. Therefore, ethical approval is not required.

Ten questions related to ear care from the Question and Answer (Q&A) section on the World Health Organisation's (WHO) official website were selected and asked to ChatGPT (16).

1. How should I clean my (or my child's) ears?
2. What is ear wax? Should it be removed?
3. When does wax need to be removed from the ear?
4. Can I use Q-tips or cotton buds to remove ear wax?
5. What causes ear pain?
6. What should I do for ear pain?
7. Can I take any antibiotic for ear pain and ear discharge?
8. What causes fluid or pus from ears? Can it be dangerous?
9. What should I do if my ears feel itchy or if an insect enters my ear?
10. What should I do when water goes into my ear?

These questions were posed in separate new chat sessions using ChatGPT's GPT-4o model on 12 June 2025 and the GPT-5 model on 15 August 2025. Default settings were used, and no custom instructions, web search functions, external tools, or additional prompts were employed. After all ten questions had been posed, the session was terminated and a new chat session was initiated. The complete set of questions was repeated seven times in separate chat sessions for each model.

The readability of ChatGPT answers was assessed using six validated, commonly used readability assessment tools: Flesch Reading Ease (FRE) score, Flesch–Kincaid grade level (FKGL), Gunning Fog Index (GFI), Coleman–Liau Index (CLI), Simple Measure of Gobbledygook (SMOG) grade, and Automated Readability Index (ARI). In 1948, Rudolph Flesch developed the Flesch Reading Ease (FRE) formula. Later, in 1975, Kincaid and colleagues built upon the FRE to create the Flesch–Kincaid Grade Level (FKGL) formula for the U.S. Navy, providing a grade-level indicator for written material (17). FRE is calculated from the total number of words, syllables, and sentences in a written passage using the following formula: 206.835 – 1.015 (total words/total sentences) – 84.6 (total syllables/total words). The score obtained from FRE ranges from 0 to 100, and the higher the score, the easier the text is to read (18). FKGL, on the other hand, provides a score corresponding to the US grade level (17). The Gunning Fog Index is a formula used to estimate the number of years of formal education required to understand a text on first reading. It is based on the average number of words per sentence and the percentage of complex words in the text. Higher scores indicate greater reading difficulty (19). The SMOG grade is calculated by comparing the number of polysyllabic words (three or more syllables) to the number of sentences. This provides an estimate of the number of years of education required for the average person to understand a given piece of writing (20). The Coleman-Liau Index is a measure that takes into account the average number of characters per word and the average number of sentences per 100 words in a given text sample. This index determines the grade level at which a text is written using a specific formula (21). The Automated Readability Index (ARI) estimates the grade level needed to understand a text on first reading, using the average number of characters per word and words per sentence (19).

Readability scores were calculated by copying ChatGPT's responses to each question into a free online readability calculator (22). Statistical analyses were conducted using the Statistical Package for the Social Sciences (SPSS) version 15.0. The normality of the data was assessed using the Kolmogorov–Smirnov and Shapiro–Wilk tests. Descriptive statistics are presented as mean ± standard deviation and median (minimum–maximum). The independent samples t-test was used for comparisons of parametric data, while the Mann–Whitney U test was applied for non-parametric data. Effect sizes were calculated as r ($Z/\sqrt{N}$) for the Mann–Whitney U test and as Cohen's d for independent-samples t-tests. A p-value<0.05 was considered statistically significant.

RESULTS

The mean FRE score was 72.93 ± 7.69 for ChatGPT-5, corresponding to the 'fairly easy' category, while it was 62.84 ± 8.32 for ChatGPT-4, which corresponds to the 'standard' level (p<0.001). The mean FKGL score was 5.75 ± 1.17 for ChatGPT-5, which is equivalent to an elementary reading level, whereas the mean score for ChatGPT-4o was 8.13 ± 1.59, corresponding to a middle school level. Across indices, the reading levels of ChatGPT-5 responses ranged from approximately 5th to 7th grade (FKGL:5.75±1.17, SMOG:6.35±0.79, GFI:7.37±1.22, CLI:7.32±1.52, ARI:5.78±1.42), while those of ChatGPT-4 responses ranged from approximately 7th to 9th grade (FKGL:8.13±1.59, SMOG:7.35±1.42, GFI:8.44±1.79, CLI:9.62±1.71, ARI:8.70±2.02). ChatGPT-5 showed better readability than ChatGPT-4 across all readability indices (p<0.001). The effect sizes ranged from 0.31 to 1.68, indicating moderate to large differences between ChatGPT-4o and ChatGPT-5. Readability scores for ChatGPT-4o and ChatGPT-5 responses are shown in Table 1.

The mean FRE score for the questions answered by the World Health Organization was 76.9 ± 8.44, with the other readability indices indicating levels ranging from 5th to 7th grade (FKGL 6.36 ± 1.58, SMOG 5.6 ± 2.39, GFI 7.02 ± 1.67, CLI 6.9 ± 1.71, ARI 6.56 ± 1.64). There was no statistically significant in readability between ChatGPT-5 and WHO responses; however, ChatGPT-4o responses differed significantly from WHO responses on all readability indices and were more difficult to read. A comparison of readability indices between ChatGPT models and WHO responses is shown in Table 2.

Table 3 presents ChatGPT responses with the highest and lowest FRE scores to the questions.

Table 1. Readability scores of ChatGPT-4o and ChatGPT-5 responses.

Readability Indices	ChatGPT-4o		ChatGPT-5		p	Effect sizes
	Mean±SD	Median(min-max)	Mean±SD	Median(min-max)		
FRE	62.84±8.32	65.0 (40.0-78.0)	72.93±7.69	74.0 (52.0-86.0)	<0.001	0.55
FKGL	8.13±1.59	7.78 (5.39-11.91)	5.75±1.17	5.54 (3.57-8.97)	<0.001	0.67
ARI	8.7±2.02	8.4 (5.12-13.01)	5.78±1.42	5.58 (2.86-9.61)	<0.001*	1.68
GFI	8.44±1.79	8.0 (6.0-13.1)	7.37±1.22	7.15 (5.6-10.6)	<0.001	0.31
CLI	9.62±1.71	9.13 (7.0-13.59)	7.32±1.52	7.12 (4.61-11.65)	<0.001	0.38
SMOG	7.35±1.42	6.95 (5.02-10.85)	6.35±0.79	6.24 (5.05-8.67)	<0.001	0.61

*Independent-samples t-test. All other comparisons were performed using the Mann–Whitney U test.

Table 2. Comparison of readability indices between ChatGPT models and WHO responses.

Readability Indices	ChatGPT-5 vs. WHO responses		ChatGPT-4o vs. WHO responses	
	p	Effect sizes	p	Effect sizes
FRE	0.104	0.18	<0.001	0.44
FKGL	0.150*	0.49	0.002	0.35
ARI	0.115*	0.54	0.002*	1.08
GFI	0.288	0.12	0.013	0.28
CLI	0.428*	0.27	<0.001	0.41
SMOG	0.171	0.15	0.011	0.28

*Independent-samples t-test. All other comparisons were performed using the Mann–Whitney U test.

Table 3. Examples of questions with the lowest and highest scores according to FRE.

Model	The question with the easiest readability based on FRE score	The question with the most difficult readability based on FRE score
ChatGPT-4o	**Can I use Q-tips or cotton buds to remove ear wax?** (FRE:78.0) **Answer:** No, it's best not to use Q-tips or cotton buds inside your ear canal. Here's why: • They can push the wax deeper into the ear, causing blockages or even damage. • They might injure the sensitive skin inside the ear canal or the eardrum. • Using them too often can cause irritation or infections. If you want to clean your ears, just gently wipe the outer ear with a soft, damp cloth. For wax buildup deeper inside, see a healthcare professional for safe removal. Would you like tips on safer ways to manage ear wax at home?	**Can I take any antibiotic for ear pain and ear discharge?** (FRE:40.0) **Answer:** No, you should not take any antibiotic for ear pain or ear discharge without a doctor's advice. Here's why: • Ear pain and discharge can have different causes, and not all of them need antibiotics. • Many ear infections are viral, which don't respond to antibiotics. • Using antibiotics unnecessarily can cause side effects and contribute to antibiotic resistance. • A healthcare provider needs to examine your ear to determine if you have a bacterial infection that requires antibiotics, and which antibiotic is appropriate. If you have ear pain with discharge, fever, or worsening symptoms, see a doctor promptly for proper diagnosis and treatment.
ChatGPT-5	**What should I do when water goes into my ear?** (FRE:86.0) **Answer:** Getting water trapped in the ear is usually harmless but can sometimes cause discomfort or infection. Here's what to do: 1. Remove the water safely • Tilt your head to the side and gently pull on the earlobe to help the water drain. • Jump or shake your head gently while tilted to encourage drainage. • Lie on your side with the affected ear facing down for a few minutes. 2. Dry the ear • Use a soft towel to gently dry the outer ear. • Do not insert cotton swabs or fingers into the ear canal. 3. Optional techniques • Gravity + jaw movement: Chewing or yawning can help water escape. • Alcohol-vinegar drops (for adults, not children under 1 year): A few drops can help dry the ear and prevent infection—but only if the eardrum is intact. 4. Watch for signs of infection ("swimmer's ear") • Pain, redness, swelling • Discharge or foul odor • Itching or fever If any of these appear, see a doctor promptly. I can also make a quick home routine to prevent water-related ear problems, especially for kids who swim or bathe often. Do you want me to do that?	**What causes fluid or pus from ears? Can it be dangerous?** (FRE:52.0) **Answer:** Fluid or pus from the ear (otorrhea) can happen for several reasons — some mild, some potentially serious. 1. Common causes • Outer ear infection (otitis externa / swimmer's ear) – infection of the ear canal, often after swimming or scratching. • Middle ear infection (otitis media) with eardrum perforation – pus drains out when pressure builds up behind the eardrum and it ruptures. • Chronic suppurative otitis media – long-term ear infection with ongoing discharge. • Fungal infection – discharge may look white, black, or fluffy. • Ear injury – from foreign objects, trauma, or sudden pressure changes. 2. Less common but serious causes • Cholesteatoma – abnormal skin growth in the middle ear that can erode bone. • Mastoiditis – infection of the bone behind the ear (often after untreated ear infection). • Cerebrospinal fluid (CSF) leak – rare, usually after head injury or surgery; fluid may be clear and watery. 3. When it can be dangerous Seek urgent medical care if discharge is: • Accompanied by fever, severe pain, swelling, or redness around the ear • Mixed with blood • Clear and watery after head trauma (possible CSF leak) • Persistent for more than a few days despite basic care • Associated with hearing loss, dizziness, or facial weakness 4. Key points • Never put anything in the ear unless prescribed by a doctor. • Keep the ear dry until evaluated. • The correct treatment (antibiotics, antifungals, surgery) depends on the exact cause — guessing can delay proper care. If you want, I can make you a visual chart showing ear discharge types, colors, and their likely causes — it's very useful for quick recognition, especially for parents.

DISCUSSION

The internet has become a major platform for individuals seeking health-related information, with more than half of the US population relying on it as a primary source (23,24). Recent advances in artificial intelligence have further expanded these opportunities, providing patients with access to a new and rapidly growing source of medical information (25). Among these tools, ChatGPT—an AI-based language model that has gained substantial attention—has the potential to become a widely used resource for health information. Therefore, ensuring the reliability and readability of its responses is of critical importance. To date, numerous studies have evaluated the reliability of ChatGPT responses. It has generally been reported to provide accurate and trustworthy information on topics such as asthma, spinal cord injuries, thyroid cancer, and hypothyroidism in pregnancy (26-29). In addition, ChatGPT has demonstrated success in medical exams (8,30). While these findings highlight its potential as a reliable source of medical knowledge, the readability of its responses is equally important, as information that is difficult to comprehend may undermine its practical value and contribute to negative health outcomes. Therefore, our study focused on assessing the readability of ChatGPT's responses regarding ear care.

Readability is a key component of effective health communication. When health information is presented at a level that is too complex, patients' ability to understand, retain, and apply the information may be limited, ultimately reducing the effectiveness of health education efforts. Therefore, assessing the readability of responses provided by large language models such as ChatGPT is essential to ensure that AI-supported health information remains accessible and understandable to a wide audience. Patients who are unable to fully comprehend medical information may be disadvantaged as active participants in their own healthcare decisions, which can jeopardize their access to the highest quality of care (24). The American Medical Association recommends that patient education materials be written at a sixth-grade reading level, while the National Institutes of Health recommends an eighth-grade level (31).

Among readability measures, the Flesch Reading Ease (FRE) is the most commonly used to evaluate the readability of written health information materials (17). Based on the FRE scores, the most understandable response was at the 'fairly easy' level for ChatGPT-4o and at the 'easy' level for ChatGPT-5 (78.0 and 86.0, respectively). The mean FRE score in our study was 62,84±8,32 for ChatGPT-4o and 72.93±7.69 for ChatGPT-5; indicating that responses generated by ChatGPT-4o are "standard" and ChatGPT-5 are 'fairly easy'. Similarly, the mean FKGL score was 8.13±1.59 for ChatGPT-4o (corresponding to an eighth-grade level) and 5.75±1.17 for ChatGPT-5 (corresponding to an elementary level). Studies suggest that patient comprehension may be compromised if content exceeds a 7th-grade reading level, which is the average American reading level identified by the United States Department of Health and Human Services (32). Taken together, our findings suggest that ChatGPT-5 provides health-related information in a more accessible and comprehensible manner compared with ChatGPT-4o, which may, in turn, facilitate better understanding and application of medical information by readers.

Our findings indicate that, when comparing ChatGPT with the WHO, ChatGPT-5 responses were closer to the readability levels of WHO materials than responses from ChatGPT-4o. WHO and ChatGPT-5 responses corresponded to a fifth- to seventh-grade reading level, which aligns more closely with recommended standards. In contrast, ChatGPT-4o responses were consistently higher, ranging from 7th- to 9th-grade levels, indicating a more complex readability profile. This comparison highlights that responses from ChatGPT-5 closely match the readability of WHO materials, indicating a notable improvement over ChatGPT-4o in producing health information accessible to the general public. However, these findings should be interpreted with caution, as WHO materials and ChatGPT-generated responses differ fundamentally in their development processes and intended purposes. WHO patient education materials are developed through expert review and editorial oversight, whereas ChatGPT responses are dynamically generated and may vary across interactions. Therefore, the present comparison reflects similarities in readability rather than equivalence in content quality, accuracy, reliability, or suitability for patient education.

In our study, the SMOG index was 7.35 for ChatGPT-4o and 6.35 for ChatGPT-5. In contrast, a previous study on the surgical treatment of retinal diseases reported a SMOG index above 13 for ChatGPT-4o responses, corresponding to college-level readability. This may be explained by the more complex medical content of that study and the fact that the questions were prepared for clinicians (33). Similarly, a study evaluating the readability of ChatGPT-3.5-turbo model responses on lung cancer reported considerably higher scores (GFI 13.63 ± 1.54, CLI 14.24 ± 1.90, SMOG 14.61 ± 1.45, ARI 15.04 ± 1.97) compared to those observed in our study (19). The technical nature of the question set used in that study may be associated with these high scores. The readability of ChatGPT-4's responses regarding hypothyroidism during pregnancy was also found to be difficult, often requiring a college-level education to comprehend the text; however, the questions in that study targeted both patients and clinicians, which may account for the higher readability levels (29). Evidence on ChatGPT-5 readability is still limited, but one recent study on polytrauma reported higher values (SMOG: 7.98 ± 1.12, CLI: 14.22 ± 3.24, ARI: 11.23 ± 1.86, FRE: 42.26 ± 6.08, FKGL: 11.24 ± 1.46) compared to ours (34). Overall, higher readability scores reported in studies of various diseases, compared with our findings, may be explained by our study's focus on ear care–related questions, which contain less complex medical content and are designed for the general public. This suggests that readability levels may depend not only on the AI model used but also on the complexity of medical topics and the intended audience for the questions.

This study has some limitations. First, it compared only ChatGPT-4o and ChatGPT-5, excluding other large language models or previous model versions. Second, the analysis was limited to English-language responses; therefore, the results may not be generalizable to responses in other languages. Although each response was generated in a separate chat session, repeated outputs from the same prompt may not be completely independent. Therefore, the effective sample size may be smaller than the nominal number of responses analysed; the findings should be interpreted accordingly. Future studies may benefit from applying repeated-measures or mixed-effects modelling approaches to account for potential within-question correlations. Another limitation is that responses from GPT-4o and GPT-5 were collected on different dates. Changes in the underlying model infrastructure, system-level settings, or platform updates occurring between the two data collection periods may have influenced the generated responses. Therefore, the observed differences cannot be attributed exclusively to the model versions.

This study demonstrated that ChatGPT-5 provides health-related responses with greater readability and closer alignment to the recommended readability standards for patient education materials than ChatGPT-4o. These findings suggest that ChatGPT-5 may enhance the general public's understanding of medical information and potentially support health education efforts. This study addresses a clear gap in the literature by evaluating the readability of ChatGPT-5 responses for which there is currently limited evidence. A direct comparison of ChatGPT-4o and ChatGPT-5 provides insight into the progress of large language models and their potential to generate more understandable health-related content. That ChatGPT-5 produces more readable outputs than ChatGPT-4o highlights ongoing advancements in artificial intelligence and suggests its growing potential as an accessible source of health information. However, the readability of ChatGPT outputs may vary depending on the complexity of the medical topic and the target audience. Since such responses may directly influence individuals' health-related decisions, presenting the information clearly and comprehensibly is essential. In our study, the use of questions designed for the general public increases the relevance of the findings to real-world health communication and patient education. Importantly, readability reflects only the ease with which information can be read and understood and does not necessarily indicate accuracy, completeness, reliability, or clinical appropriateness. Therefore, the superior readability observed for ChatGPT-5 should not be interpreted as evidence of greater medical accuracy or overall information quality. Future studies should evaluate these additional dimensions to provide a more comprehensive assessment of large language models in health communication. Future development of ChatGPT should focus on tailoring responses to users' educational backgrounds to maximize comprehension and ensure safe, equitable access to health information.

CONCLUSION

In conclusion, ChatGPT-5 generated significantly more readable responses to ear care questions than ChatGPT-4o, with readability levels closer to those of WHO patient education materials. These findings suggest that newer large language models may have greater potential for providing accessible health information to the general public. However, readability alone does not reflect the accuracy, completeness, or clinical appropriateness of AI-generated information, and further studies evaluating these dimensions are warranted.

Ethical approval

In this study, since no patient or experimental animal data were used, ethical committee approval was not required.

Author contribution

Concept: CÖ; Design: CÖ, MFK; Data Collection or Processing: CÖ, MFK; Analysis or Interpretation: CÖ, MFK; Literature Search: CÖ, MFK; Writing: CÖ. All authors reviewed the results and approved the final version of the article.

Source of funding

The authors declare the study received no funding.

Conflict of interest

The authors declare that there is no conflict of interest.

REFERENCES

1. Koirala B, Manandhar S, Shah RK. Knowledge, attitude and practice of ear care among pre clinical medical students at Birat Medical College Teaching Hospital. Birat Journal of Health Sciences. 2020; 5(3): 1236-1240. **[Crossref]**
2. Sağlam Ö, Saat B, Dursun E, et al. The prevalence of cerumen in young male population. KBB-Forum. 2013; 12(2): 32-5.
3. Alateeq OM. Knowledge Attitude and practice of ear care in Saudi community. Baqai J Health Sci. 2017; 20(2): 11-4.
4. Alrajhi MS, Alim BM, Aldokhayel SD, Zeitouni LM, Al Tawil LK, Alzahrani FA. Knowledge, attitudes, and practices pertaining to cotton-bud usages and the complications related to their misuse among outpatients in an ear, nose, and throat clinic. Journal of Nature and Science of Medicine. 2019; 2(4): 220-5.
5. Tonsaker T, Bartlett G, Trpkov C. Health information on the Internet: gold mine or minefield? Can Fam Physician. 2014; 60(5): 407-8.
6. Beheshti M, Toubal IE, Alaboud K, et al. Evaluating the reliability of chatgpt for health-related questions: a systematic review. Informatics. 2025; 12(1): 9. **[Crossref]**
7. Johnson D, Goodman R, Patrinely J, et al. Assessing the accuracy and reliability of AI-generated medical responses: An evaluation of the Chat-GPT model. Research Square [Preprint]. 2023. **[Crossref]**
8. Kung TH, Cheatham M, Medenilla A, et al. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. PLOS Digit Health. 2023; 2(2): e0000198. **[Crossref]**
9. Antaki F, Touma S, Milad D, El-Khoury J, Duval R. Evaluating the performance of ChatGPT in ophthalmology: An analysis of its successes and shortcomings. Ophthalmol Sci. 2023; 3(4): 100324. **[Crossref]**
10. Huh S. Are ChatGPT's knowledge and interpretation ability comparable to those of medical students in Korea for taking a parasitology examination?: a descriptive study. J Educ Eval Health Prof. 2023; 20: 1. **[Crossref]**
11. Wang S, Mo C, Chen Y, Dai X, Wang H, Shen X. Exploring the performance of ChatGPT-4 in the Taiwan audiologist qualification examination: Preliminary observational study highlighting the potential of AI chatbots in hearing care. JMIR Med Educ. 2024; 10: e55595. **[Crossref]**
12. Zalzal HG, Abraham A, Cheng J, Shah RK. Can ChatGPT help patients answer their otolaryngology questions? Laryngoscope Investig Otolaryngol. 2024; 9(1): e1193. **[Crossref]**
13. Qu RW, Qureshi U, Petersen G, Lee SC. Diagnostic and management applications of ChatGPT in structured otolaryngology clinical scenarios. OTO Open. 2023; 7(3): e67. **[Crossref]**
14. Soon S, Perry B. Paging Dr. ChatGPT: safety, accuracy and readability of ChatGPT in ENT emergencies. Australian Journal of Otolaryngology. 2025; 8: 8. **[Crossref]**
15. Adams RC, Smith TP, Metts JK, Ross JW. Readability-its applicability to education of patients by pharmacy. Hosp Pharm. 1979; 14(11): 654-6.
16. World Health Organization (WHO). Deafness and hearing loss: ear care. Available at: https://www.who.int/news-room/questions-and-answers/item/deafness-and-hearing-loss-ear-care (Accessed on June 4, 2025).
17. Jindal P, MacDermid JC. Assessing reading levels of health information: uses and limitations of flesch formula. Educ Health (Abingdon). 2017; 30(1): 84-8. **[Crossref]**

18. Flesch R. A new readability yardstick. J Appl Psychol. 1948; 32(3): 221-33. **[Crossref]**

19. Gencer A. Readability analysis of ChatGPT's responses on lung cancer. Sci Rep. 2024; 14(1): 17234. **[Crossref]**

20. Hedman AS. Using the SMOG formula to revise a health-related document. Am J Health Educ. 2008; 39(1): 61-4. **[Crossref]**

21. Azer SA, AlOlayan TI, AlGhamdi MA, AlSanea MA. Inflammatory bowel disease: An evaluation of health information on the internet. World J Gastroenterol. 2017; 23(9): 1676-96. **[Crossref]**

22. Readability Scoring System. Available at: https://readabilityformulas.com/readability-scoring-system.php (Accessed on August 20, 2025).

23. Baumann E, Czerwinski F, Rosset M, Seelig M, Suhr R. How do people in Germany seek health information? Insights from the first wave of HINTS Germany. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz. 2020; 63(9): 1151-60. **[Crossref]**

24. Kianian R, Carter M, Finkelshtein I, Eleswarapu SV, Kachroo N. Application of artificial intelligence to patient-targeted health information on kidney stone disease. J Ren Nutr. 2024; 34(2): 170-6. **[Crossref]**

25. Fritsch SJ, Blankenheim A, Wahl A, et al. Attitudes and perception of artificial intelligence in healthcare: A cross-sectional survey among patients. Digit Health. 2022; 8: 20552076221116772. **[Crossref]**

26. Høj S, Thomsen SF, Ulrik CS, Meteran H, Sigsgaard T, Meteran H. Evaluating the scientific reliability of ChatGPT as a source of information on asthma. J Allergy Clin Immunol Glob. 2024; 3(4): 100330. **[Crossref]**

27. Özcan F, Örücü Atar M, Köroğlu Ö, Yılmaz B. Assessment of the reliability and usability of ChatGPT in response to spinal cord injury questions. J Spinal Cord Med. 2025; 48(5): 852-7. **[Crossref]**

28. Cavnar Helvaci B, Hepsen S, Candemir B, et al. Assessing the accuracy and reliability of ChatGPT's medical responses about thyroid cancer. Int J Med Inform. 2024; 191: 105593. **[Crossref]**

29. Onder CE, Koc G, Gokbulut P, Taskaldiran I, Kuskonmaz SM. Evaluation of the reliability and readability of ChatGPT-4 responses regarding hypothyroidism during pregnancy. Sci Rep. 2024; 14(1): 243. **[Crossref]**

30. Long C, Lowe K, Santos A, et al. Evaluating ChatGPT-4 in otolaryngology-head and neck surgery board examination using the CVSA model. medRxiv [Preprint]. 2023. **[Crossref]**

31. Rooney MK, Santiago G, Perni S, et al. Readability of Patient Education Materials From High-Impact Medical Journals: A 20-Year Analysis. J Patient Exp. 2021; 8: 2374373521998847. **[Crossref]**

32. Walsh TM, Volsko TA. Readability assessment of internet-based consumer health information. Respir Care. 2008; 53(10): 1310-15. **[Crossref]**

33. Momenaei B, Wakabayashi T, Shahlaee A, et al. Appropriateness and readability of ChatGPT-4-generated responses for surgical treatment of retinal diseases. Ophthalmol Retina. 2023; 7(10): 862-8. **[Crossref]**

34. Şimşek SD, Şensöğüt C. The use of readability formulas in polytrauma. Acta Medica Young Doctors. 2025; 1(1): 2-9. **[Crossref]**