

Original Paper

Assessing the Readability of Online Patient Education Materials in Obstetrics and Gynecology Using Traditional Measures: Comparative Analysis and Limitations

Anunita Nattam^{1*}, BSc; Tripura Vithala^{1*}; Tzu-Chun Wu¹, PhD; Shwetha Bindhu¹, BSc; Gregory Bond², MSc; Hexuan Liu³, PhD; Amy Thompson¹, MD; Danny T Y Wu^{1,2}, MSI, PhD

¹College of Medicine, University of Cincinnati, Cincinnati, OH, United States

²Department of Computer Science, University of Cincinnati, Cincinnati, OH, United States

³School of Criminal Justice, University of Cincinnati, Cincinnati, OH, United States

*these authors contributed equally

Corresponding Author:

Danny T Y Wu, MSI, PhD

Department of Computer Science

University of Cincinnati

2610 University Circle

Cincinnati, OH, 45221

United States

Phone: 1 5135564307

Email: wutz@ucmail.uc.edu

Abstract

Background: Patient education materials (PEMs) can be vital sources of information for the general population. However, despite American Medical Association (AMA) and National Institutes of Health (NIH) recommendations to make PEMs easier to read for patients with low health literacy, they often do not adhere to these recommendations. The readability of online PEMs in the obstetrics and gynecology (OB/GYN) field, in particular, has not been thoroughly investigated.

Objective: The study sampled online OB/GYN PEMs and aimed to examine (1) agreeability across traditional readability measures (TRMs), (2) adherence of online PEMs to AMA and NIH recommendations, and (3) whether the readability level of online PEMs varied by web-based source and medical topic. This study is not a scoping review, rather, it focused on scoring the readability of OB/GYN PEMs using the traditional measures to add empirical evidence to the literature.

Methods: A total of 1576 online OB/GYN PEMs were collected via 3 major search engines. In total 93 were excluded due to shorter content (less than 100 words), yielding 1483 PEMs for analysis. Each PEM was scored by 4 TRMs, including Flesch-Kincaid grade level, Gunning fog index, Simple Measure of Gobbledygook, and the Dale-Chall. The PEMs were categorized based on publication source and medical topic by 2 research team members. The readability scores of the categories were compared statistically.

Results: Results indicated that the 4 TRMs did not agree with each other, leading to the use of an averaged readability (composite) score for comparison. The composite scores across all online PEMs were not normally distributed and had a median at the 11th grade. Governmental PEMs were the easiest to read amongst source categorizations and PEMs about menstruation were the most difficult to read. However, the differences in the readability scores among the sources and the topics were small.

Conclusions: This study found that online OB/GYN PEMs did not meet the AMA and NIH readability recommendations and would be difficult to read and comprehend for patients with low health literacy. Both findings connected well to the literature. This study highlights the need to improve the readability of OB/GYN PEMs to help patients make informed decisions. Research has been done to create more sophisticated readability measures for medical and health documents. Once validated, these tools need to be used by web-based content creators of health education materials.

(*J Med Internet Res* 2023;25:e46346) doi: [10.2196/46346](https://doi.org/10.2196/46346)

KEYWORDS

obstetrics and gynecology; online patient education materials; readability; online education; education; health literacy; literature; medical documents; tool; utilization; online content; health education; education material; assessment; obstetrics; gynecology

Introduction

Readability assessment of online health information (OHI) provides valuable insight into how much information is easily understandable, to effectively empower laypersons to make informed health decisions [1]. Specifically, recognizing the impact of the situational reading level demands of the population and implementing that knowledge to simplify OHI, particularly materials geared toward patient education, can act as a mediating factor in health and clinical decision-making for patients by simplifying clinician-patient communication and reducing the complexities of the health care system for the patient [2]. The aim of online patient education materials (PEMs) should be to support interactive health literacy within an individual, giving them the ability to extract health information and derive meaning from different sources. This ability grants patients the opportunity to engage in interactions with health care professionals, fostering greater understanding and shared decision making. Therefore, using the internet population's reading level and literacy capacity as a metric to guide PEMs can increase the usability and effectiveness of PEMs [3]. Furthermore, as the internet now serves as the primary source of information in modern society, and the prevalence of smartphones has allowed the expansion of the internet to a wider population, improving the quality of OHI and PEMs can reach a diversity of populations, and can decrease the burdens of frontline professional support and patients alike [4].

Current recommendations made by the American Medical Association (AMA) and the National Institutes of Health (NIH) suggest that all PEMs should be at a sixth-eighth-grade reading level or lower [5,6]. However, previous studies conducted broadly throughout clinical specialties indicate these guidelines are not regularly followed [7]. Moreover, our preliminary scoping review conducted prior to this study showed that there is a lack of readability assessment studies in the field of obstetrics and gynecology (OB/GYN). Since OB/GYN PEMs found on the internet are often accessed to seek guidance on a wide range of symptoms, diagnoses, and treatments, it provides patients with information on a spectrum of preventable and curable gynecological diseases. Of note, 1 important public health example is human papillomavirus and cervical cancer, for which preventative care via vaccination and early detection through Papanicolaou test screening are available. Given that OB/GYN is a primary health care field, lack of access to

appropriate online PEMs may be a limiting factor in overcoming poor interactive health literacy and its associated outcomes [8].

Over recent years, popular traditional readability measures (TRMs), such as the Flesch-Kincaid grade level (FKGL) [9], Gunning fog index (GFI) [10], Simple Measure of Gobbledygook (SMOG) [11], and the Dale-Chall (DCL) [12] formulas generate a grade-level score associating readability with the grade level of education needed to understand a document. For example, a document with a readability of the 5th-grade level is easier to read and understand than a document with a 10th-grade reading level. The TRMs used similar textual features in their formula (Table 1), including the average number of words per sentence, the average number of syllables per word, the average number of sentences, and custom easy versus difficult word lists.

Previous research done to assess the readability of OB/GYN PEMs, although scarce, has shown that most existing OB/GYN PEMs are written between the 9th and 12th grade level [13-18]. However, several of these studies focus solely on PEMs published by academic sources, limiting the study to a specific subtype of PEMs that patients are likely to encounter on the internet [13,14]. Moreover, in these studies where TRMs were used, scores generated by each measure were not always consistent, indicating a limitation related to the validity of these TRMs used to widely assess patient readability and understanding [19].

Therefore, to address the gaps in existing research and further address the issue of readability in online OB/GYN PEMs, 4 research questions were produced. First, do the scores generated by 4 TRMs agree with each other (research question 1 [RQ1])? This research question was prompted by previous research indicating that certain readability measures do not agree with each other, particularly FKGL, SMOG, and GFI [20-22]. Second, do the PEMs found in the field of OB/GYN follow the sixth-eighth-grade level readability recommendation by the AMA and NIH (research question 2 [RQ2])? Third, are there differences in readability level by sources (government agencies, nonprofit organizations, educational websites, or commercial entities) (research question 3 [RQ3])? Finally, will the readability level change when discussing different gynecological processes and topics, such as menstruation, pregnancy, cancer, general disease, and procedural information (research question 4 [RQ4])?

Table 1. Mathematical equation used for each TRM.^a

Formula	Equation
Flesch-Kincaid grade level (FKGL) [9]	Reading grade level = $(0.39 \times \text{average number of words per sentence}) + (11.8 \times \text{average number of syllables per word}) - 15.59$
Gunning fog index (GFI) [10] ^b	Reading grade level = $0.4 (\text{average number of words per sentence} + \text{number of words with 3 or more syllables} \times [100 / \text{number of words}])$
Simple Measure of Gobbledygook (SMOG) [11]	<p>If number of sentences ≥ 30: reading grade level = 3 + square root of polysyllable count to the nearest perfect square.</p> <p>Else if number of sentences is between 1 and 29: raw score = average number of polysyllable per sentence \times ratio of sentences) + number of polysyllable</p> <p>Round the Raw score to the integer.</p> <p>Reading grade level</p> <p>= 5 if Raw score is between 1 and 6</p> <p>= R if Raw score is in $[R^2 - 7 \times R + 13, (R - 2) \times (R - 3)]$ where R is between 6 and 17</p> <p>= 18 if Raw score is equal to or larger than 211</p>
Dale-Chall (DCL) [12]	<p>Raw score = $(0.0496 \times \text{average number of words per sentence}) + (15.79 \times \text{number of words not found on a word list}) / (\text{number of words}) + 3.6365$</p> <p>Reading grade level</p> <p>= 4 if Raw score is smaller than 5</p> <p>= $2 \times \text{Raw score} - 5$ if Raw score is between 5 and 8</p> <p>= $3 \times \text{Raw score} - 14$ if Raw score is 9</p> <p>= 16 if Raw score is equal to or larger than 10</p>

^aTRM: traditional readability measures.

^bRequires at least 100 words in the input document.

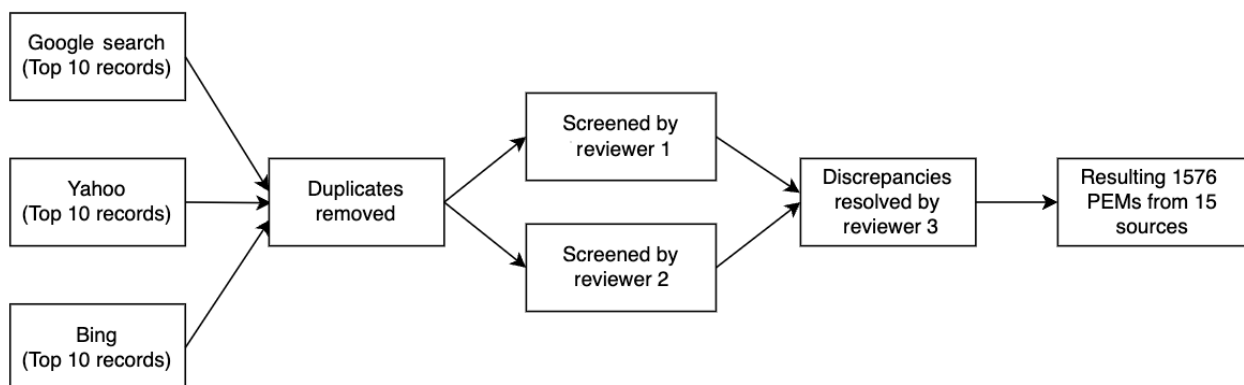
Methods

Data Collection and Coding

The PEMs were selected by searching the keywords “OB/GYN” AND “Patient Education Materials” in the 3 most used search engines (ie, Google, Yahoo, and Bing!), and selecting the results on the first page (top 10 results) from each. This is not considered as a scoping review. Patient search behavior and patterns were mimicked in order to generate the most realistic search results. After removing repetitions and broken links, 15 different website sources from varying categories were identified containing 1576 PEMs (Multimedia Appendix 1). The PEMs were crawled by using the Python Selenium library, which

automatically visited the individual web pages and extracted the full text for each paper with formatting retained. In terms of coding, 2 reviewers (TV and Somya Pandey) surveyed the texts together and coded a small sample to determine the codebook by sources and topics. This process is considered as a bottom-up approach as opposed to a top-down approach using predefined categories. Then, the 2 reviewers coded all the PEMs independently to produce the source and topic categories. Any discrepancy was resolved by the third reviewer (AN). Of note, each PEM might have multiple topics. The primary topic of each PEMs was determined and used for the subsequent statistical analysis. The process of data collection and screening is shown in Figure 1.

Figure 1. Data collection and screening process. PEM: patient education material.



Readability Scoring

Among the 1576 PEMs crawled, 93 of them did not have enough number of words for GFI scoring (minimum 100 words). These PEMs were removed in the readability scoring and subsequent analysis, resulting in a total of 1483 OB/GYN PEMs analyzed. To obtain readability scores, the text of the collected PEMs was first cleaned, primarily by removing the title and any sections unrelated to the health material, such as mentions of other papers and advertisements, and then scored based on the TRM formulas. Specifically, each PEM was broken down into sentences, words, and syllables. Sentences were derived by splitting the text by common end-of-sentence punctuation, while words were found by using a space delimiter for each sentence; syllables and base words were found mainly through the Natural Language Tool Kit Python library. The Natural Language Tool Kit Python library is a collection of modules to process text and perform natural language processing (NLP) tasks [23]. Surface metrics of the text, such as the average number of words per sentence, were then calculated based on the 4 TRMs to generate the scores. Of note, the DCL raw scores were transformed into grade levels using its reading grade level formula (Table 1). Only the integer part of the result from the mathematical equations will be preserved as the reading grade level. A random sample of 50 PEMs was selected and hand calculations were conducted and compared to the Python-generated values to ensure validity of our methods. Of note, the decision to use these 4 TRMs (FKGL, GFI, SMOG, and DC) was based on their use in previous studies assessing the readability of online PEMs, although not particularly in the field of OB/GYN [24-27].

Data Analysis

To address our research questions, respective hypotheses were proposed. First (RQ1), it was hypothesized that the TRMs would have acceptable agreeability (hypothesis 1) since they used similar textual features and determined readability as grade levels. To test this hypothesis, each readability score of a document was converted to a grade level from grade 4 to 16 due to the grade level mapping scale of DCL and treated as categorical data. In addition, the study tried to improve the agreement by combining the categories, for example, less than sixth grade, sixth-eighth-grade, and above eighth grade. The Fleiss κ was calculated to measure the degree of agreement between the different readability measures over that which would be expected by chance [28]. The given κ value can be a negative number up to 1.00, with negative numbers indicating poor agreement, 0.00-0.20 as slight agreement, 0.21-0.40 as fair agreement, 0.41-0.60 as moderate agreement, 0.61-0.80 as

substantial agreement, and 0.81-1.00 being almost perfect to perfect agreement. When hypothesis 1 is held, meaning all 4 TRMs agree with each other substantially, FKGL will be used to represent the readability of a document since it is widely used (eg, FKGL is implemented in Word [Microsoft Corp]). Otherwise, a composite score (average or median of all 4 TRMs) will be created to represent the readability of a document.

Second (RQ2), we hypothesized that the PEMs would not follow the recommended guidelines of sixth-eighth-grade level readability (hypothesis 2) since literature of readability assessment has shown this trend. To test this hypothesis, the composite scores were used to generate both categorical and numerical data. The normality of the composite scores was examined using the Shapiro-Wilk test. The null hypothesis of this test is to assume that the population of data is normally distributed. Since the significant value of the Shapiro-Wilk test on our data was less than 0.05, the population of data deviates significantly from a normal distribution. Therefore, Wilcoxon Sign test was used to examine if the composite scores overall are above the recommended level (eighth grade).

Lastly, the study explored the possibility of differences in readability scores for PEMs coming from varying sources (RQ3) and topics (RQ4). It was hypothesized that the government sourced PEMs would have the lowest readability level (hypothesis 3), making them the easiest to understand, since the government agencies claimed that their online PEMs have been curated and are suitable for the public. On the other hand, it was hypothesized that the readability level of PEMs with different topics was similar (hypothesis 4). To test these 2 hypotheses, the PEMs were manually categorized into 4 source categories (ie, government, commercial, nonprofit, and educational) and 5 topic categories (general disease, pregnancy, menstruation, procedure, and cancer). General disease referred to any papers referencing diseases that do not affect women exclusively, cannot be placed into any of the other categories, and appeared on these OB/GYN PEM websites. The Kruskal-Wallis test was applied to detect any differences between medians in each category (main effect), followed by the pairwise Mann-Whitney test with Bonferroni correction as the post hoc test to assess whether the median of the composite scores of the different categories were statistically different (Figure 2). The median of the composite score was used because the composite scores were not normally distributed. Table 2 shows distribution of the PEMs analyzed in this study by the source and topic categorization.

Figure 2. Process of data analysis. FKGL: Flesch-Kincaid grade level; HSD: honestly significant difference; PEM: patient education material; TRM: traditional readability measure.

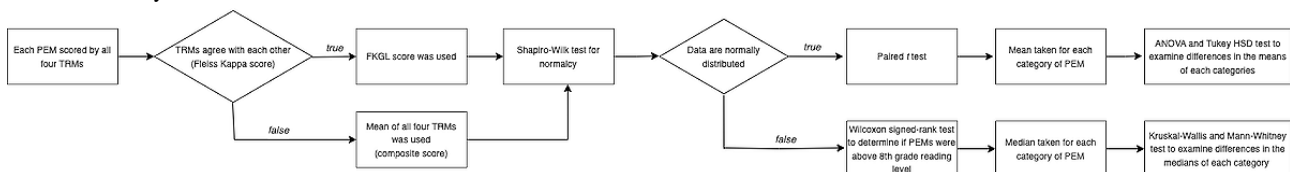


Table 2. Contingency table of online OB/GYN PEMs^a sources based on topics.

Topic or source	Educational, n	Nonprofit, n	Commercial, n	Government, n	Total, n
Pregnancy	553	531	34	27	1145
General disease	39	13	57	28	137
Procedure	42	22	24	10	98
Menstruation	24	3	22	6	55
Cancer	4	1	31	12	48
Total	662	570	168	83	1483

^aOB/GYN PEM: obstetrics and gynecology patient education material.

Ethical Considerations

Of note, the present study did not require IRB review since it was based on publicly available records with no identifiers of individuals.

Results

Hypothesis 1

Placing the PEMs into 13 categories from 4 to 16, that is 1 per grade level, resulted in a very low Fleiss $\kappa=0.0025$ value, which indicates slight agreement. Placing the readability scores into 3 categories (less than sixth grade, sixth to eighth grade, and greater than eighth grade) only resulted in a slight increase in the Fleiss κ ($P=.033$) and placing them into just 2 categories (less than or equal to 8 and greater than 8) resulted in a significant increase ($P=.08$). However, the agreement is still

considered slight. As a result, the 4 TRMs did not perform fair agreement in our data set. Then, the composite score of a document was generated by averaging the 4 TRMs. [Figure 1](#) shows the distribution of the composite scores using both the mean and median. As shown in [Figure 1](#), taking the average of the readability scores can be more conservative since the average considered a variety of responses in each TRM. On the other hand, the median involved only 2 out of 4 responses in our case, which resulted in the loss of half the information.

Hypothesis 2

Using the composite score of the TRMs, the PEMs were found to be at the 11th grade reading level, which is significantly higher than the recommended 8th grade level ($P<.001$) using the Wilcoxon sign test. The median only was used in these calculations, as the mean was shown to be skewed by extreme values ([Table 3](#)).

Table 3. Distribution of the composite scores (average of the 4 TRMs^a; RQ2^b).

Composite readability grade level	PEMs ^c , n (%)	Accumulated number, n (%)
6	1 (0.06)	1 (0.06)
7	20 (1.27)	21 (1.33)
8	44 (2.79)	65 (4.12)
9	100 (6.35)	165 (10.47)
10	178 (11.29)	343 (21.76)
11	206 (13.07)	549 (34.84)
12	321 (20.37)	870 (55.20)
13	276 (17.51)	1146 (72.72)
14	197 (12.50)	1343 (85.22)
15	183 (11.61)	1526 (96.83)
16	50 (3.17)	1576 (100)

^aTRM: traditional readability measures.

^bRQ2: research question 2.

^cPEM: patient education material.

Hypothesis 3

The Kruskal-Wallis test showed that at least 1 median is significantly different from others ($P<.001$) among the 4 documentation sources. The pairwise Mann-Whitney test indicated that the median readability level of the government

source (11.25) was significantly lower than the commercial (13.00), educational (12.75), and nonprofit (12.75). Meanwhile, the median readability level of the commercial source (13.00) was significantly higher than that of the nonprofit source (12.75), but not different from the educational source (12.75). None of

the categories had a median readability level less than the recommended eighth-grade level (Table 4).

Table 4. Results from pairwise Mann-Whitney tests^a (RQ3^b).

Source category comparison		Median of mean		Mann-Whitney statistic	P value
Category 1	Category 2	Median 1	Median 2		
Commercial	Government	13.00	11.25	11,223	<.01 ^a
Commercial	Educational	13.00	12.75	62,131	.02
Commercial	Nonprofit	13.00	12.75	54,775	.004 ^a
Government	Educational	11.25	12.75	16,356	<.01
Government	Nonprofit	11.25	12.75	15,189	<.01 ^a
Educational	Nonprofit	12.75	12.75	192,935	.46

^a $P < .008$.

^bRQ3: research question 3.

Hypothesis 4

The Kruskal-Wallis test showed there is at least 1 median which is significantly different from others ($P = .001$) among the 5 medical topics. The pairwise Mann-Whitney test indicated that

the median readability levels of many groups were similar to each other. However, the median readability level of menstruation (13.5) and pregnancy (12.75) are significantly higher than general disease (12.25; Table 5).

Table 5. Results from pairwise Mann-Whitney tests^a (RQ4^b).

Topic category comparison		Composite scores		Mann-Whitney statistic	P value
Category 1	Category 2	Median 1	Median 2		
Menstruation	Procedure	13.5	12.75	3369	.01
Menstruation	General disease	13.5	12.25	5271.5	<.01 ^a
Menstruation	Cancer	13.5	12.5	1720.5	.02
Menstruation	Pregnancy	13.5	12.75	38,180	.007
Procedure	General disease	12.75	12.25	7886.5	.02
Procedure	Cancer	12.75	12.5	2422.5	.93
Procedure	Pregnancy	12.75	12.75	54,094	.56
General disease	Cancer	12.25	12.5	2747.5	.06
General disease	Pregnancy	12.25	12.75	62,984	<.01 ^a
Cancer	Pregnancy	12.5	12.75	26,927	.64

^a $P < .005$.

^bRQ4: research question 4.

Discussion

Principal Findings

This study applied 4 TRMs to score the readability of OB/GYN online PEMs and compared the scores of the TRMs, which is the first to conduct such comparative analysis and provide empirical evidence to show the limitations of the TRMs. The findings showed that the 4 TRMs lacked agreement with one another, indicating the need for developing modern readability measures for PEMs and general OHI. Using the average score of TRMs as a proxy to the readability level, our study found that most OB/GYN PEMs required high readability levels despite the recommendation for online PEMs to be at the

sixth-eighth-grade reading level. In addition, the study found that the online PEMs from the government source were slightly easier to read than other sources. However, they still require 11th grade level. It is important to provide health information in a consistent language so that OHI can be disseminated to all, regardless of source. Lastly, the comparison of the topic categories showed that the menstruation and pregnancy PEMs are harder to read than the general disease. This opens new research questions and research investigations.

Implications

The finding where the 4 TRMs did not agree with one another connects well with recent research concerning TRMs. For example, 1 study evaluating the readability of diabetes-related

PEMs found that certain readability measures consistently indicated higher levels of readability than others [22]. This indicates that the existing measures do not reflect completely the readability level of health texts [29]. Moving away from syntax and surface language features, recent research has identified other factors, such as sentence complexity, use of passive voice, grammar frequency, and the patient's familiarity with vocabulary, as additional components that need to be considered in measuring health text readability [30-32]. Passive voice, in particular, was shown to distinguish very complex texts from very simple texts, but was not able to distinguish intermediate complexities, making it an important language feature to study further [33]. Additionally, the evaluation of readability measures is critical, especially using human annotations since the readability measures may not reflect the actual readability and understandability of a piece of text. Previous studies have found such discrepancy between the perceived document difficulty and actual difficulty [19,34]. Our study highlights the need to improve the readability of OB/GYN PEMs and ensure the validity and reliability of readability measures.

When looking at readability assessment studies done in medical specialties other than OB/GYN, required readability levels are consistently very high [35-37], showing a need for remediation across specialties like ophthalmology, cardiology, geriatric care, and beyond. For example, in 1 study specifically using TRM to analyze ophthalmology related PEMs, average readability was found to be at the 10th-grade reading level [35]. Even with inconsistent TRM levels, human annotations have rated PEMs in several medical fields to be too difficult to understand [38]. Such studies have been repeated in various specialties over several years, highlighting the issue of readability in health education as a pertinent issue for some time.

Over the years, some studies have moved away from the use of TRM, most likely due to their poor ability to accurately assess health information. Recently developed readability assessment methods have adopted machine learning, artificial intelligence, and NLP techniques, which have been shown to better predict readability when compared to human annotations [29,39,40]. However, these methods have not been widely used or promoted in the research community. With further advancement of artificial intelligence and NLP, such as pretrained large language models, it may become easier to assess readability, and even revise and generate easier-to-read health information automatically.

Limitations

This study has at least 6 limitations. First, the PEM search may not be comprehensive; it may miss other online PEMs in the field of OB/GYN. However, we used modern search engines to collect PEMs that were most likely to be accessed by

laypersons. Therefore, the PEMs included in this study should be frequently seen by laypersons. Since the study only included the results on the first page of each search engine, relevant websites may be listed on the second page or after. Second, the PEMs were scored by only 4 TRMs. There are other TRMs (such as Fry readability graph [41] and FORCAST [42]) that could potentially be more useful and better adept at identifying the true readability of online PEMs. Moreover, there have been other readability measures developed recently for health information [19,29,31]. However, since many of the new readability assessment methods have not been widely used by the research community, they were excluded in this study. Third, in this study, the texts for all PEMs were taken without any content analysis, which can limit our understanding of the actual readability level. Fourth, although statistical significance was found in the results, there is not sufficient evidence to prove that small differences in grade level above grade 12 are meaningful since most readability measures are calibrated for school grades. Next, although the PEMs were cleaned, the title and some sections of the PEMs were removed, which may affect the readability score distribution [43]. Lastly, the search results might have been slightly altered due to the presence of cache and other confounders in Google search [44].

Future Work

Future work includes surveying the literature and summarizing the current advances in readability formula development, given the fact that although research has been conducted to create new readability formulas, very few are promoted and widely used to replace TRMs in the research community. Implementing these health information specific readability measures may provide some more insight into what must be done to improve the readability of online PEMs. Additionally, there will be further content analysis conducted on PEMs to create writing guidelines to support clinicians and patients alike, especially in the context of OB/GYN. Patient stakeholders or standardized patients can also inform the creation of new PEM content for validation of health care-specific readability tools. Lastly, a simple English thesaurus for OB/GYN PEMs can be developed to help those writing PEMs to simplify their materials.

Conclusions

This study examined the readability of the PEMs in OB/GYN. While the OB/GYN PEMs were collected using search engines, which may introduce biases, the study found that most of the PEMs were hard to read, requiring high school or college level of education to read the content. More research is needed to evaluate the PEMs in OB/GYN in a more comprehensive manner and create writing guidelines to improve the readability, understandability, and actionability of these PEMs.

Acknowledgments

We would like to thank Mr Mostafa Ghaderian for his effort in statistical analysis. We also thank Ms Somya Pandey for her effort in document categorization and copyediting.

Conflicts of Interest

None declared.

Multimedia Appendix 1

The PEM website source information and categorization.

[\[DOCX File, 25 KB-Multimedia Appendix 1\]](#)

References

1. Edmunds MR, Barry RJ, Denniston AK. Readability assessment of online ophthalmic patient information. *JAMA Ophthalmol* 2013;131(12):1610-1616 [doi: [10.1001/jamaophthalmol.2013.5521](https://doi.org/10.1001/jamaophthalmol.2013.5521)] [Medline: [24178035](https://pubmed.ncbi.nlm.nih.gov/24178035/)]
2. Elwyn G, Laitner S, Coulter A, Walker E, Watson P, Thomson R. Implementing shared decision making in the NHS. *BMJ* 2010;341:c5146 [FREE Full text] [doi: [10.1136/bmj.c5146](https://doi.org/10.1136/bmj.c5146)] [Medline: [20947577](https://pubmed.ncbi.nlm.nih.gov/20947577/)]
3. Nutbeam D, Lloyd JE. Understanding and responding to health literacy as a social determinant of health. *Annu Rev Public Health* 2021;42:159-173 [FREE Full text] [doi: [10.1146/annurev-publhealth-090419-102529](https://doi.org/10.1146/annurev-publhealth-090419-102529)] [Medline: [33035427](https://pubmed.ncbi.nlm.nih.gov/33035427/)]
4. Friis K, Lasgaard M, Osborne RH, Maindal HT. Gaps in understanding health and engagement with healthcare providers across common long-term conditions: a population survey of health literacy in 29,473 Danish citizens. *BMJ Open* 2016;6(1):e009627 [FREE Full text] [doi: [10.1136/bmjopen-2015-009627](https://doi.org/10.1136/bmjopen-2015-009627)] [Medline: [26769783](https://pubmed.ncbi.nlm.nih.gov/26769783/)]
5. Abu-Heija A, Shatta M, Ajam M, Abu-Heija U, Imran N, Levine D. Quantitative readability assessment of the internal medicine online patient information on Annals.org. *Cureus* 2019;11(3):e4184 [FREE Full text] [doi: [10.7759/cureus.4184](https://doi.org/10.7759/cureus.4184)] [Medline: [31106084](https://pubmed.ncbi.nlm.nih.gov/31106084/)]
6. Weiss BD. Health literacy: a manual for clinicians. American Medical Association. 2003. URL: <http://lib.ncfh.org/pdfs/6617.pdf> [accessed 2023-06-05]
7. Rooney MK, Santiago G, Perni S, Horowitz DP, McCall AR, Einstein AJ, et al. Readability of patient education materials from high-impact medical journals: a 20-year analysis. *J Patient Exp* 2021;8:2374373521998847 [FREE Full text] [doi: [10.1177/2374373521998847](https://doi.org/10.1177/2374373521998847)] [Medline: [34179407](https://pubmed.ncbi.nlm.nih.gov/34179407/)]
8. Cudjoe J, Budhathoki C, Roter D, Gallo JJ, Sharps P, Han-Ra H. Exploring health literacy and the correlates of Pap testing among African immigrant women: findings from the AfroPap study. *J Cancer Educ* 2021;36(3):441-451 [FREE Full text] [doi: [10.1007/s13187-020-01755-9](https://doi.org/10.1007/s13187-020-01755-9)] [Medline: [32410109](https://pubmed.ncbi.nlm.nih.gov/32410109/)]
9. Flesch R. A new readability yardstick. *J Appl Psychol* 1948;32(3):221-233 [doi: [10.1037/h0057532](https://doi.org/10.1037/h0057532)] [Medline: [18867058](https://pubmed.ncbi.nlm.nih.gov/18867058/)]
10. Gunning R. *The Technique of Clear Writing*. New York: McGraw-Hill; 1968.
11. Mc Laughlin GH. SMOG grading-a new readability formula. *J Read* 1969;12(8):639-646
12. Chall JS, Dale E. *Manual for Use of the New Dale-Chall Readability Formula*. Cambridge, MA: Brookline Books; 1995.
13. Fahimuddin FZ, Sidhu S, Agrawal A. Reading level of online patient education materials from major obstetrics and gynecology societies. *Obstet Gynecol* 2019;133(5):987-993 [doi: [10.1097/AOG.0000000000003214](https://doi.org/10.1097/AOG.0000000000003214)] [Medline: [30969212](https://pubmed.ncbi.nlm.nih.gov/30969212/)]
14. Okuhara T, Okada H, Goto E, Kiuchi T. Readability assessment of HPV vaccination and cervical cancer information: a systematic scoping review. *Healthcare (Basel)* 2021;9(10):1246 [FREE Full text] [doi: [10.3390/healthcare9101246](https://doi.org/10.3390/healthcare9101246)] [Medline: [34682926](https://pubmed.ncbi.nlm.nih.gov/34682926/)]
15. Lange EMS, Shah AM, Braithwaite BA, You WB, Wong CA, Grobman WA, et al. Readability, content, and quality of online patient education materials on preeclampsia. *Hypertens Pregnancy* 2015;34(3):383-390 [doi: [10.3109/10641955.2015.1053607](https://doi.org/10.3109/10641955.2015.1053607)] [Medline: [26153628](https://pubmed.ncbi.nlm.nih.gov/26153628/)]
16. Yee LM, Niznik CM, Simon MA. Examining the role of health literacy in optimizing the care of pregnant women with diabetes. *Am J Perinatol* 2016;33(13):1242-1249 [doi: [10.1055/s-0036-1584540](https://doi.org/10.1055/s-0036-1584540)] [Medline: [27322666](https://pubmed.ncbi.nlm.nih.gov/27322666/)]
17. AlKhalili R, Shukla PA, Patel RH, Sanghvi S, Hubbi B. Readability assessment of internet-based patient education materials related to mammography for breast cancer screening. *Acad Radiol* 2015;22(3):290-295 [doi: [10.1016/j.acra.2014.10.009](https://doi.org/10.1016/j.acra.2014.10.009)] [Medline: [25488695](https://pubmed.ncbi.nlm.nih.gov/25488695/)]
18. Freda MC, Damus K, Merkatz IR. Evaluation of the readability of ACOG patient education pamphlets. *The American College of Obstetricians and Gynecologists. Obstet Gynecol* 1999;93(5 Pt 1):771-774 [doi: [10.1016/s0029-7844\(98\)00518-3](https://doi.org/10.1016/s0029-7844(98)00518-3)] [Medline: [10912984](https://pubmed.ncbi.nlm.nih.gov/10912984/)]
19. Kauchak D, Leroy G. Moving beyond readability metrics for health-related text simplification. *IT Prof* 2016;18(3):45-51 [FREE Full text] [doi: [10.1109/MITP.2016.50](https://doi.org/10.1109/MITP.2016.50)] [Medline: [27698611](https://pubmed.ncbi.nlm.nih.gov/27698611/)]
20. Wu DTY, Hanauer DA, Mei Q, Clark PM, An LC, Proulx J, et al. Assessing the readability of ClinicalTrials.gov. *J Am Med Inform Assoc* 2016;23(2):269-275 [FREE Full text] [doi: [10.1093/jamia/ocv062](https://doi.org/10.1093/jamia/ocv062)] [Medline: [26269536](https://pubmed.ncbi.nlm.nih.gov/26269536/)]
21. Mailloux SL, Johnson ME, Fisher DG, Pettibone TJ. How reliable is computerized assessment of readability? *Comput Nurs* 1995;13(5):221-225 [Medline: [7585304](https://pubmed.ncbi.nlm.nih.gov/7585304/)]
22. Lipari M, Berlie H, Saleh Y, Hang P, Moser L. Understandability, actionability, and readability of online patient education materials about diabetes mellitus. *Am J Health Syst Pharm* 2019;76(3):182-186 [doi: [10.1093/ajhp/zxy021](https://doi.org/10.1093/ajhp/zxy021)] [Medline: [31408087](https://pubmed.ncbi.nlm.nih.gov/31408087/)]
23. Lobur M, Romanyuk A, Romanyshyn M. Using NLTK for educational and scientific purposes. : IEEE; 2011 Presented at: 2011 11th International Conference The Experience of Designing and Application of CAD Systems in Microelectronics (CADSM); 23-25 February 2011; Polyana, Ukraine p. 426-428

24. Cheng BT, Kim AB, Tanna AP. Readability of online patient education materials for glaucoma. *J Glaucoma* 2022;31(6):438-442 [doi: [10.1097/IJG.0000000000002012](https://doi.org/10.1097/IJG.0000000000002012)] [Medline: [35283441](https://pubmed.ncbi.nlm.nih.gov/35283441/)]
25. Pearson K, Ngo S, Ekpo E, Sarraju A, Baird G, Knowles J, et al. Online patient education materials related to lipoprotein(a): readability assessment. *J Med Internet Res* 2022;24(1):e31284 [FREE Full text] [doi: [10.2196/31284](https://doi.org/10.2196/31284)] [Medline: [35014955](https://pubmed.ncbi.nlm.nih.gov/35014955/)]
26. Fortuna J, Riddering A, Shuster L, Lopez-Jeng C. Assessment of online patient education materials designed for people with age-related macular degeneration. *BMC Ophthalmol* 2020;20(1):391 [FREE Full text] [doi: [10.1186/s12886-020-01664-x](https://doi.org/10.1186/s12886-020-01664-x)] [Medline: [33008367](https://pubmed.ncbi.nlm.nih.gov/33008367/)]
27. Martin CA, Khan S, Lee R, Do AT, Sridhar J, Crowell EL, et al. Readability and suitability of online patient education materials for glaucoma. *Ophthalmol Glaucoma* 2022;5(5):525-530 [doi: [10.1016/j.ogla.2022.03.004](https://doi.org/10.1016/j.ogla.2022.03.004)] [Medline: [35301989](https://pubmed.ncbi.nlm.nih.gov/35301989/)]
28. Fleiss JL. Measuring nominal scale agreement among many raters. *Psychol Bull* 1971;76(5):378-382 [doi: [10.1037/h0031619](https://doi.org/10.1037/h0031619)]
29. Kim H, Goryachev S, Rosemblat G, Browne A, Keselman A, Zeng-Treitler Q. Beyond surface characteristics: a new health text-specific readability measurement. *AMIA Annu Symp Proc* 2007;2007:418-422 [FREE Full text] [Medline: [18693870](https://pubmed.ncbi.nlm.nih.gov/18693870/)]
30. Leroy G, Endicott JE, Kauchak D, Mouradi O, Just M. User evaluation of the effects of a text simplification algorithm using term familiarity on perception, understanding, learning, and information retention. *J Med Internet Res* 2013;15(7):e144 [FREE Full text] [doi: [10.2196/jmir.2569](https://doi.org/10.2196/jmir.2569)] [Medline: [23903235](https://pubmed.ncbi.nlm.nih.gov/23903235/)]
31. Zheng J, Yu H. Assessing the readability of medical documents: a ranking approach. *JMIR Med Inform* 2018;6(1):e17 [FREE Full text] [doi: [10.2196/medinform.8611](https://doi.org/10.2196/medinform.8611)] [Medline: [29572199](https://pubmed.ncbi.nlm.nih.gov/29572199/)]
32. Ji M, Liu Y, Hao T. Predicting health material accessibility: development of machine learning algorithms. *JMIR Med Inform* 2021;9(9):e29175 [FREE Full text] [doi: [10.2196/29175](https://doi.org/10.2196/29175)] [Medline: [34468321](https://pubmed.ncbi.nlm.nih.gov/34468321/)]
33. Ownby RL. Influence of vocabulary and sentence complexity and passive voice on the readability of consumer-oriented mental health information on the internet. *AMIA Annu Symp Proc* 2005;2005:585-589 [FREE Full text] [Medline: [16779107](https://pubmed.ncbi.nlm.nih.gov/16779107/)]
34. Kauchak D, Leroy G, Hogue A. Measuring text difficulty using parse-tree frequency. *J Assoc Inf Sci Technol* 2017;68(9):2088-2100 [FREE Full text] [doi: [10.1002/asi.23855](https://doi.org/10.1002/asi.23855)] [Medline: [29057293](https://pubmed.ncbi.nlm.nih.gov/29057293/)]
35. Williams AM, Muir KW, Rosdahl JA. Readability of patient education materials in ophthalmology: a single-institution study and systematic review. *BMC Ophthalmol* 2016;16:133 [FREE Full text] [doi: [10.1186/s12886-016-0315-0](https://doi.org/10.1186/s12886-016-0315-0)] [Medline: [27487960](https://pubmed.ncbi.nlm.nih.gov/27487960/)]
36. Avra TD, Le M, Hernandez S, Thure K, Ulloa JG. Readability assessment of online peripheral artery disease education materials. *J Vasc Surg* 2022;76(6):1728-1732 [doi: [10.1016/j.jvs.2022.07.022](https://doi.org/10.1016/j.jvs.2022.07.022)] [Medline: [35931399](https://pubmed.ncbi.nlm.nih.gov/35931399/)]
37. Wang E, Kalloniatis M, Ly A. Assessment of patient education materials for age-related macular degeneration. *Ophthalmic Physiol Opt* 2022;42(4):839-848 [FREE Full text] [doi: [10.1111/opo.12991](https://doi.org/10.1111/opo.12991)] [Medline: [35521818](https://pubmed.ncbi.nlm.nih.gov/35521818/)]
38. Alshehri A, Alghofaili N, Alshunaiber R, Alkadi L. Quality and readability assessment of internet-based information on common prosthodontic treatments. *Int J Prosthodont* 2022;35(1):62-67 [doi: [10.11607/ijp.7063](https://doi.org/10.11607/ijp.7063)] [Medline: [33651024](https://pubmed.ncbi.nlm.nih.gov/33651024/)]
39. Sung YT, Chen JL, Cha JH, Tseng HC, Chang TH, Chang KE. Constructing and validating readability models: the method of integrating multilevel linguistic features with machine learning. *Behav Res Methods* 2015;47(2):340-354 [doi: [10.3758/s13428-014-0459-x](https://doi.org/10.3758/s13428-014-0459-x)] [Medline: [24687843](https://pubmed.ncbi.nlm.nih.gov/24687843/)]
40. Mukherjee P, Leroy G, Kauchak D. Using lexical chains to identify text difficulty: a corpus statistics and classification study. *IEEE J Biomed Health Inform* 2019;23(5):2164-2173 [FREE Full text] [doi: [10.1109/JBHI.2018.2885465](https://doi.org/10.1109/JBHI.2018.2885465)] [Medline: [30530380](https://pubmed.ncbi.nlm.nih.gov/30530380/)]
41. Fry E. A readability formula that saves time. *J Read* 1968;11(7):513-578
42. Caylor JS, Sticht TG, Fox LC, Ford JP. Methodologies for determining reading requirements of military occupational specialties. ERIC. 1973. URL: <https://files.eric.ed.gov/fulltext/ED074343.pdf> [accessed 2023-07-18]
43. Mac O, Ayre J, Bell K, McCaffery K, Muscat DM. Comparison of readability scores for written health information across formulas using automated vs manual measures. *JAMA Netw Open* 2022;5(12):e2246051 [FREE Full text] [doi: [10.1001/jamanetworkopen.2022.46051](https://doi.org/10.1001/jamanetworkopen.2022.46051)] [Medline: [36508219](https://pubmed.ncbi.nlm.nih.gov/36508219/)]
44. Fajardo MA, Weir KR, Bonner C, Gnjdjic D, Jansen J. Availability and readability of patient education materials for deprescribing: an environmental scan. *Br J Clin Pharmacol* 2019;85(7):1396-1406 [FREE Full text] [doi: [10.1111/bcp.13912](https://doi.org/10.1111/bcp.13912)] [Medline: [30848837](https://pubmed.ncbi.nlm.nih.gov/30848837/)]

Abbreviations

- AMA:** American Medical Association
- DCL:** Dale-Chall
- FKGL:** Flesch-Kincaid grade level
- GFI:** Gunning fog index
- NIH:** National Institutes of Health
- NLP:** natural language processing
- OB/GYN:** obstetrics and gynecology
- OHI:** online health information
- PEM:** patient education material

SMOG: Simple Measure of Gobbledygook

RQ1: research question 1

RQ2: research question 2

RQ3: research question 3

RQ4: research question 4

TRM: traditional readability measures

Edited by T Leung, T de Azevedo Cardoso; submitted 07.02.23; peer-reviewed by J Ayre, J Zheng, C Jacob, CI Sartorão Filho; comments to author 18.04.23; revised version received 06.06.23; accepted 04.07.23; published 30.08.23

Please cite as:

Nattam A, Vithala T, Wu TC, Bindhu S, Bond G, Liu H, Thompson A, Wu DTY

Assessing the Readability of Online Patient Education Materials in Obstetrics and Gynecology Using Traditional Measures: Comparative Analysis and Limitations

J Med Internet Res 2023;25:e46346

URL: <https://www.jmir.org/2023/1/e46346>

doi: [10.2196/46346](https://doi.org/10.2196/46346)

PMID: [37647115](https://pubmed.ncbi.nlm.nih.gov/37647115/)

©Anunita Nattam, Tripura Vithala, Tzu-Chun Wu, Shwetha Bindhu, Gregory Bond, Hexuan Liu, Amy Thompson, Danny T Y Wu. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 30.08.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.