



Using the Patient Education Materials Assessment Tool (PEMAT) to assess Parkinson's Disease informational content websites

Ashok P¹, Lim J.G², Bapuraj J.R³, Masotti M⁴, Spears C.C⁵

Submitted: September 15, 2024, Revised: version 1, November 8, 2024, version 2, November 14, 2024, version 3, November 26, 2024

Accepted: December 2, 2024

Abstract

This study performed an assessment of Parkinson's disease (PD) informational content websites using the grade level reading method and the Patient Education Materials Assessment Tool (PEMAT). Google and Bing searches were performed on 2/27/2023 and 3/3/2023 respectively for PD. A website's grade level readability was determined using Flesch reading ease, Flesch reading level, and passive sentences percentage scores. The website informational content was also evaluated using the PEMAT A/V test for understandability and actionability by two readers, a board-certified radiologist and a high school student. Websites (n=105) were categorized into academic (30), governmental (17), health media (31), medical product companies (4), non-academic (5) and professional societies (18). Overall median scores for Flesch Reading Ease, Reading level and Passive sentence % were 40.3, 11.9 and 15.6% respectively. Median PEMAT understandability and actionability scores for reader 1 and 2 were {91%, 100%} and {90%, 100%} respectively with no significant differences between website categories. Interobserver agreement for PEMAT was significant ($k=0.787$). While the average grade reading level presented at the websites was higher than that recommended for the public, the acceptable PEMAT understandability and actionability scores suggested that the online information was appropriate from a patient education standpoint. The actionability and understandability of online informational PD related content; as measured by the PEMAT; may not necessarily be compromised even if Grade level reading scores reflect otherwise. The PEMAT hence captures the targeted utilitarian and pedagogical gestalt of informational websites, as opposed to the Grade level reading score, which only records the extraneous mechanics of grammar and reading ease. Also, the excellent interobserver agreement between readers of significantly different educational levels demonstrated the robustness of PEMAT.

Keywords

Patient Education Materials Assessment Tool, PEMAT, Parkinson's disease, Understandability, Actionability, Interobserver agreement, Grade level scoring system, Informational website

¹Corresponding author: Parth Ashok, Greenhills School, 850 Greenhills Dr., Ann Arbor, MI 48105, USA.

pashok@greenhillsschool.org

²Jerald G. Lim, Institute of Radiology, St. Luke's Medical Center, 5th Ave, Taguig, 1634 Metro Manila, Philippines.

³Jayapalli R Bapuraj, Department of Radiology, University of Michigan, 1500 E Medical Center Dr. Ann Arbor, MI 48109, USA.

⁴Maria Masotti, Department of Biostatistics, University of Michigan, 1415 Washington Heights, Ann Arbor, MI 48109, USA.
mmasotti@umich.edu

⁵Christopher C. Spears, Department of Neurology, University of Michigan, Neurology Clinic Taubman Center, 1500 E Medical Center Dr, Ann Arbor, MI 48109, USA.

Introduction

Parkinson's disease (PD), a common neurological disorder affecting more than 10 million people worldwide, and nearly a million people in the United States, has seen a significant increase in new cases diagnosed annually (1). With the increasing impact of digital resources and their availability, patients and their caregivers commonly rely on the internet for further information and next steps, therefore it becomes imperative for websites to present information that can be easily understood with potential action items that can be subsequently performed. Health agencies like the Centers for Disease Control and Prevention (CDC), the National Institutes of Health (NIH), and American Medical Association (AMA), suggest that medical information for the public be written at no higher than an eighth-grade reading level (2).

In 2010, Fitzsimmons et al. investigated PD websites with the Flesch-Kincaid and Simple Measure Of Gobbledygook (SMOG) formulae and concluded that a very small percentage (<1%) of the top 100 PD information webpages were fully comprehensible to the average adult (3). This paper is however more than a decade old and there are no recent studies assessing the readability of electronic websites dedicated to PD.

Furthermore, while readability is one aspect of this process, the grade level scoring system does not capture all the nuances associated with understandability. To address this, the Patient Education Materials Assessment Tool (PEMAT) scoring system was developed with tests for audio/visual materials (A/V) and printable materials (P) (4). The tool consists of

score, and an actionability score. This tool is designed as a guide to help determine whether patients or their providers are able to understand and act on the presented information at the website. The PEMAT contains multiple questions related to understandability and actionability which are answered by a reviewer after reading the website educational content. A composite score is then generated from the understandability and actionability domains.

There are currently no published studies analyzing PD websites using PEMAT. Therefore, the objective of our study was to comprehensively assess PD informational websites using both the grade level reading method and the more nuanced PEMAT tool. Since PEMAT is a relatively new tool which may incorporate subjectivity components, we also analyzed the interobserver agreement using observers with varying levels of educational backgrounds.

Methods

Using Google and Bing search engines, an initial list of websites related to PD was generated. The Google and Bing searches was performed on 2/27/2023 and 3/3/2023 respectively. According to <http://statista.com>, Google and Bing had the two highest market shares in Feb 2023, of 85.64%, and 8.18% respectively, hence these were chosen as the two main search engines for the study. Many search terms were considered, such as "parkinson's disease", "parkinson disease", "parkinsons", etc., but eventually the term "parkinson's disease" was used, since "parkinson disease" and "parkinsons" were autocorrected to "parkinson's disease". On both web browsers, the search term was

entered in a guest account, to ensure that past searches did not affect the results.

The top 150 search results were downloaded from each search engine. Sites requiring payment/subscription, duplicate sites, non-functional sites, those irrelevant to patients, journal/news articles, and promotional advertisements were excluded. Websites were categorized into those created by academic medical centers, non-academic medical centers, governmental agencies, health media, medical product companies, and professional societies.

Readability was determined using Microsoft Word 2023 and three parameters were generated; Flesch reading ease, Flesch reading level, and passive sentences percentage. Flesch

reading ease scores above 60 and passive sentences percentage score below 10% are considered to be easier to understand for the general reader (5,6). The Flesch reading level score corresponds to the grade level for reading and ideally would be eighth grade level or lower for better understandability. For each website, these parameters were recorded on a spreadsheet (see supplementary file). Only the main website content was analyzed and peripheral information including related papers, related articles, quick links, promoted articles, about Dr XYZ etc. were excluded from the analysis. Table 1 shows examples of websites belonging to their respective categories, along with the 3 readability parameters generated by Microsoft word.

Table 1. Representative examples of reading scores across websites of different categories are depicted. Flesch Reading Ease, Flesch Reading Levels, and Passive Sentence % are shown for examples of the five website categories. These scores are based on the website content accessed via Google search on 2/27/2023 and were calculated using Microsoft Word 2023.

Website category	Link	Flesch Reading Ease	Flesch Reading Level	Passive Sentences (%)
Academic Medical Center	https://umiamihealth.org/en/treatments-and-services/neurology/parkinson%E2%80%99s-disease	35.2	12.8	23.5
Governmental	https://dph.illinois.gov/topics-services/diseases-and-conditions/diseases-a-z-list/parkinsons.html	39.2	11.5	22.2
Health Media	https://www.brainandlife.org/disorders-a-z/parkinsons-disease/	41.0	11.3	16.6
Medical Product Company / Misc	https://curascriptsd.com/Newsroom/what-is-parkinsons-disease	35.4	12.3	22.2
Non-academic Medical Center	https://neurologyassociatesva.com/warning-signs-of-parkinsons-disease-and-when-to-see-a-neurologist/	58.0	9.6	9.5
Organization / Society	https://getpalliativecare.org/whatis/disease-types/parkinsons-disease-palliative-care/	51.7	10.7	15.9

For the next part of the project, the PEMAT audio-visual tool was used to assess each website for understandability and actionability by a high school (P.A.) student. Notably, PEMAT contains scoring systems for printable

material and audio/visual material, but does not explicitly mention websites. Hence, the Agency for Healthcare Research and Quality (AHRQ) was contacted, and it was confirmed that the audio/visual materials test is appropriate for

website analysis. The PEMAT tool can be accessed at <https://www.ahrq.gov/health-literacy/patient-education/pemat-av.html>

To assess variation between observers for this tool, a board certified radiologist (J.L.) also performed the same analysis on the websites included in the study. The training levels of the two observers (P.A. and J.L.) were intentionally designed to be significantly different in order to assess the generalizability of the PEMAT tool across individuals with widely varying educational backgrounds. Hence, the PEMAT analysis was performed by 2 raters.

Statistical analysis

Continuous data (such as Flesch Reading Ease, Reading level, Passive sentences percentage, PEMAT understandability and actionability scores for readers 1 and 2) were reported as median and interquartile range. Interobserver agreement for PEMAT was studied using intraclass correlation coefficient with overall agreement studied using the kappa test. The differences between various website categories for Flesch Reading Ease, Reading level, Passive sentences percentage, PEMAT understandability and actionability were

assessed using non-parametric tests due to the small sample size of some of the categories. A two-tailed p value < 0.05 was considered statistically significant. R Version 4.4.1 was utilized for all statistical analysis.

Results

After exclusion criteria were applied to the 150 top searches, a total of 105 websites were included in the study and analyzed (Figure 1); the distribution being 30 academic, 17 governmental, 31 health media, 4 medical product companies, 5 non-academic and 18 professional societies.

The overall median scores and interquartile ranges for Flesch Reading Ease, Reading level and Passive sentence % for each category of websites is shown in Table 2 and in Figure 2. PEMAT understandability and actionability scores (representing an average of the two readers) for various categories of websites are also shown in Figure 2. Notably, among all categories, the websites from medical product companies had the lowest Flesch reading ease scores, highest reading levels, lowest PEMAT actionability scores and understandability scores, although these differences were not statistically significant (Table 2).

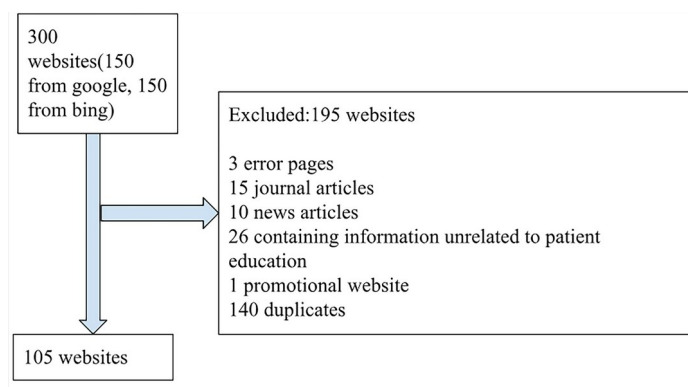


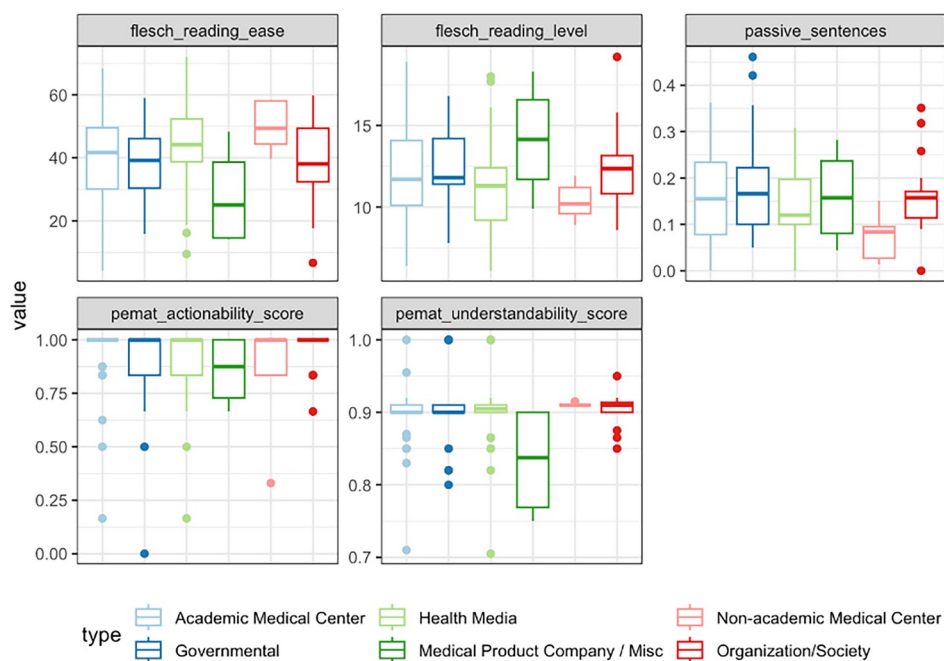
Figure 1. Flowchart demonstrating the selection process for websites included in the study.

TABLE 2. Median and Interquartile Range (IQR) for each metric by type.

Characteristic	Academic Medical Center, N = 30 ¹	Governmental, N = 17 ¹	Health Media, N = 31 ¹	Medical Product Company / Misc, N = 4 ¹	Non-academic Medical Center, N = 5 ¹	Organization/Society, N = 18 ¹	p-value ²
flesch_reading_ease	42 (30, 50)	39 (30, 46)	44 (39, 52)	25 (15, 39)	49 (44, 58)	38 (32, 49)	0.11
flesch_reading_level	11.70 (10.10, 14.08)	11.80 (11.40, 14.20)	11.30 (9.20, 12.40)	14.15 (11.70, 16.58)	10.20 (9.60, 11.20)	12.35 (10.83, 13.15)	0.2
passive_sentences	0.16 (0.08, 0.23)	0.17 (0.10, 0.22)	0.12 (0.10, 0.20)	0.16 (0.08, 0.24)	0.08 (0.03, 0.10)	0.16 (0.11, 0.17)	0.2
pemat_understandability_score	0.90 (0.90, 0.91)	0.90 (0.90, 0.91)	0.91 (0.90, 0.91)	0.84 (0.77, 0.90)	0.91 (0.91, 0.91)	0.91 (0.90, 0.91)	0.2
pemat_actionability_score	1.00 (1.00, 1.00)	1.00 (0.84, 1.00)	1.00 (0.84, 1.00)	0.88 (0.73, 1.00)	1.00 (0.84, 1.00)	1.00 (1.00, 1.00)	0.8

Averaged score of two readers is presented. ¹ Median (IQR); n (%)² Kruskal-Wallis rank sum test (continuous) or Fisher's exact test (categorical).

Distribution of scores by type of website (averaged over readers)

**Figure 2.** Distribution of scores by type of website (averaged over readers).

Interobserver agreement

The overall median PEMAT understandability and actionability scores for both readers were high and were 91% and 100% for reader 1 and 90% and 100% for reader 2 respectively.

Out of the 17 questions in the PEMAT analysis, the only questions where there was less than 100% agreement were numbers 1, 5, 13, 20, 21, 22; the least agreement was for question 22 which showed an 83.3% agreement (Q: *Does the material break down any action into manageable, explicit steps?*). The Intraclass correlation coefficient for understandability and actionability was 0.561

[CI-0.377 to 0.695] and 0.496 [CI-0.328 to 0.632] respectively, both of which are considered to be of moderate reliability. It is interesting to note that for understandability, reader 1 (high school student) had overall higher scores than reader 2 (medical professional) (Figure 3A), although no such trend was noticed for actionability (Figure 3B). We also found that actionability and understandability were not significantly correlated overall (correlation = 0.09, $p = 0.174$). Interobserver agreement for overall PEMAT assessment as an entire set was substantial ($k=0.787$).

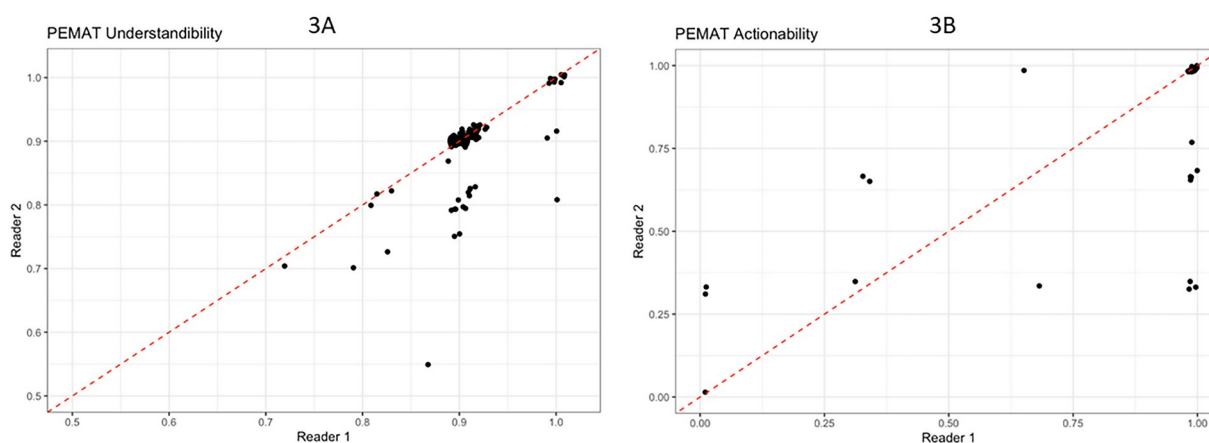


Figure 3. Correlation scatterplots demonstrating interobserver agreement between readers 1 and 2 for PEMAT understandability (A) and actionability (B). The agreement on understandability (A) is moderate (ICC 0.561) with higher scores given by reader 1 (high school student) relative to reader 2 (medical professional). For actionability (B), agreement is moderate (ICC 0.496), although no trend is seen between the two readers. The aggregate data for all 105 websites are depicted in these figures.

Discussion

Medical information for the public should be written at no higher than an eighth-grade reading level to best serve the intended audience (2). Keeping the grade level low is important, as research has found that a person's reading level is usually 3 to 5 years below the

highest grade that the person has completed. However, reading level as determined using formulae is often over simplistic and relies on average length of words and sentences (and not on the familiarity or comprehension of these words). A website can also be enhanced to be further understood by the addition of visual

aids and action items, which assist readers with information on what they can do about their medical issue. Although prior studies have assessed websites using traditional methods, one significant difference between this study and the existing readability studies for PD is the addition of the PEMAT assessment since this provides a more nuanced method to analyze websites, specifically for areas related to actionability. Also, the current study provides a more updated analysis compared to a similar single study performed more than a decade ago.

Our results showed that educational material was presented at a reading level above 8th grade for all categories of websites and there were no statistically significant differences among these categories. This is similar to prior results in the literature (4). However, when the more nuanced PEMAT tool was used, the median understandability and actionability scores were high ($\geq 90\%$), implying very good understandability and actionability. When the sites of different categories were compared to each other, no significant differences between the categories of websites were found, which likely implied that websites created by non-academic centers were just as understandable as websites from categories like those created by professional societies. The PEMAT tool also showed moderate reliability when two readers with significantly different training used it, suggesting that it was a robust evaluation tool.

The uniqueness and strength of our paper lies in the fact that one of the readers is a high school student. Most studies on inter-rater reliability of PEMAT have been performed by health professionals, who have a background

knowledge of the medical condition. Hence, while the PEMAT is designed to be used by general public and health professionals, the perspective of the general public is difficult to determine based on existing literature (7).

Limitations

One of the limitations of this study is that this was performed on two search engines, with only 150 search terms in both. Also, search results can vary depending on the geographical location and prior search history. However, since Google and Bing have the highest market search with $> 90\%$ of searches, our data likely captures what the vast majority of patients are likely to encounter in their searches.

An important caveat when interpreting PEMAT scores is that the tool is primarily meant for assessing understandability and actionability of the material. It does not check for accuracy of information. Therefore, a high PEMAT score simply means that it is understandable but does not guarantee accurate information. Additionally, although PEMAT encourages and emphasizes the use of active voice (item 5), it does not provide specific guidance on the proportion or percentage of statements across the entire webpage that should be in active voice.

Another limitation of PEMAT scores could be the binary nature of some of the questions. For instance, in the actionability assessment, although most sites identified at least one action, and hence scored high in this domain, the details on performing the action and next steps were not always clearly defined. Finally, while PEMAT is an important tool for all individuals involved with patient education, it

is not necessarily designed to determine the user-friendliness of a website.

Conclusion

While the average grade reading level was higher than recommended for the general public for PD related websites, the above average PEMAT understandability and actionability scores suggested that the information presented was appropriate from a patient education viewpoint. The actionability and understandability of online informational PD related content; as measured by the PEMAT; may not necessarily be compromised even if Grade level reading scores (as measured

by the Flesch reading ease, Flesch reading level, and passive sentences percentage scores) reflect otherwise. The PEMAT hence captures the targeted utilitarian and pedagogical gestalt of informational websites, as opposed to the Grade level reading score, which only records the extraneous mechanics of grammar and reading ease. This conclusion is corroborated by the fact that there is above average interobserver agreement between PEMAT scores for readers of substantially different educational levels, thereby further magnifying the disconnect between the largely irrelevant Grade level reading score method and the robust PEMAT tool.

References

1. Statistics. Parkinson's Foundation. <https://www.parkinson.org/understanding-parkinsons/statistics>.
2. Mishra V, Dexter JP. Comparison of Readability of Official Public Health Information About COVID-19 on Websites of International Agencies and the Governments of 15 Countries. JAMA Netw Open. 3(8), e2018033, 2020. <https://doi.org/10.1001/jamanetworkopen.2020.18033>
3. Fitzsimmons PR, Michael BD, Hulley JL, Scott GO. A readability assessment of online Parkinson's disease information. J R Coll Physicians Edinb. 40(4), 292-6, 2010. <https://doi.org/10.4997/jrcpe.2010.401>
4. The Patient Education Materials Assessment Tool (PEMAT) and user's guide. AHRQ. <https://www.ahrq.gov/health-literacy/patient-education/pemat.html>.
5. Flesch, Rudolf, How to Write in Plain English, A Book for Lawyers and Consumers, 1979, Harper and Rowe, New York, NY. pp. 82-93. <https://www.amazon.com/Write-Plain-English-Before-After/dp/0060112786>
6. Kincaid, J.P., Fishburne, R.P., Rogers, R.L., Chissom, B.S. Derivation of new readability formulas (automated readability index, fog count, and flesch reading ease formula) for Navy enlisted personnel. Research Branch Report 8-75, 1975. Chief of Naval Technical Training: Naval Air Station Memphis) <https://stars.library.ucf.edu/cgi/viewcontent.cgi?article=1055&context=istlibrary>

7. Vishnevetsky J, Walters CB, Tan KS. Interrater reliability of the Patient Education Materials Assessment Tool (PEMAT). *Patient Educ Couns*. 101(3), 490-6, 2018.
<https://doi.org/10.1016/j.pec.2017.09.003>