

Peer-Review

Nguyen, Sienna L. 2026. "Readability of Online Melanoma Patient Education Materials." *Journal of High School Science* 10 (3): 667–84. <https://doi.org/10.64336/001c.171663>.

This manuscript addresses an important topic in patient health literacy and is methodologically sound. However, the work is largely descriptive and would benefit from a more balanced discussion of its methodological limitations and the scope of its conclusions. I encourage the authors to revise the manuscript by addressing the following concerns.

1.Limited novelty

The overall study design closely follows the methodology employed in numerous previous readability studies: identifying webpages, calculating established readability indices, and comparing the results against recommended reading levels. While applying this framework to melanoma patient education materials is useful, the manuscript should more clearly acknowledge that its contribution is primarily an updated application of an established methodology rather than the introduction of a novel analytical framework. The Discussion should better position the work within the existing literature and explicitly describe its incremental contribution.

STRONG RECOMMENDATION

What would elevate the novelty:

Given the conceptual overlap among readability formulas, the authors should consider performing principal component analysis (PCA) to determine whether the various indices measure a single latent construct or multiple independent dimensions of textual complexity. If multiple principal components emerge, examining the factor loadings could identify which linguistic characteristics (e.g., sentence length, word complexity, character density, or multisyllabic word usage) dominate each component. Such findings would extend the manuscript beyond a descriptive readability survey by providing quantitative guidance for optimizing future patient education materials. Rather than attempting to improve every readability index simultaneously, authors of future educational webpages could target the linguistic features associated with the principal components that explain the greatest proportion of the variance, thereby developing materials that are optimized across multiple dimensions of readability.

An additional analysis could project each webpage into the principal component space, allowing the identification of outlier resources that achieve favorable readability despite covering complex clinical material. Examining the writing characteristics of these high-performing webpages may reveal practical strategies for producing accessible patient education materials without sacrificing scientific accuracy.

This is considerably more interesting scientifically because it shifts the question from: "Are melanoma webpages difficult to read?" to

"What independent dimensions define readability, and how can future webpages be systematically optimized along those dimensions?"

That transforms the paper from a confirmatory survey into a study that begins to generate a framework for evidence-based design of patient education materials. If the PCA reveals two or three dominant components explaining most of the variance, future authors could aim to optimize their text with respect to those components rather than treating six different readability indices as separate optimization targets. This would provide a principled, data-driven approach to developing more accessible educational content.

2.Limited representativeness of the sample

The study analyzes 36 webpages from 18 institutions, which is adequate for evaluating major institutional melanoma patient education resources. However, the manuscript frequently generalizes its findings to online melanoma patient education materials as a whole. The online ecosystem contains hundreds of educational resources from academic centers, professional organizations, commercial providers, advocacy groups, and international organizations. Accordingly, the authors should clarify that their conclusions are restricted to the sampled institutional resources and may not fully represent the broader landscape of online melanoma educational materials.

3.Arbitrary website selection

The rationale for selecting only the first 18 Google search results should be better justified. The manuscript does not explain why this number was chosen or whether the conclusions would remain unchanged if additional websites were included. Furthermore, Google search rankings are dynamic and may vary according to geographic location, personalization, browser history, and changes in search algorithms. The Discussion should explicitly acknowledge these potential sources of sampling bias and explain that the objective was to evaluate highly visible institutional resources rather than exhaustively characterize all available websites.

4.Website rankings change over time

Because Google search results evolve continuously, the study may not be fully reproducible if repeated at a later date. Different investigators conducting the same search may retrieve a different set of webpages due to search engine updates or user-specific factors. The manuscript should acknowledge this limitation and clarify that the analyzed webpages represent a snapshot of the online environment at the time the search was performed rather than a permanent ranking of melanoma educational resources.

5.Readability is only one component of accessibility

The manuscript repeatedly associates readability with accessibility, although readability represents only one dimension of patient education quality. Overall accessibility is also influenced by page organization, illustrations, graphics, multimedia content, cultural appropriateness, navigation, language availability, and other usability factors. The Discussion should acknowledge that readability formulas quantify textual complexity only and therefore should not be interpreted as comprehensive measures of accessibility.

6.No assessment of information quality

The present study evaluates readability but does not assess the quality, completeness, or accuracy of the educational content itself. Readability alone does not determine whether patient education materials effectively communicate accurate and clinically useful information. Established instruments such as DISCERN, PEMAT, JAMA benchmark criteria, or measures of understandability and actionability could provide complementary evaluations of educational quality.

7.No direct assessment of patient comprehension

Although the manuscript appropriately notes that readability formulas do not directly measure patient comprehension, the conclusions should be tempered accordingly. The study demonstrates that readability scores exceed recommended levels but does not establish that patients actually experience reduced understanding, poorer decision-making, or worse clinical outcomes as a consequence. Future work involving patient testing would be required to establish these relationships, and this distinction should be emphasized more clearly throughout the Discussion and Conclusion.

8.Limited statistical analysis

The statistical analyses are restricted primarily to descriptive statistics, t-tests, and ANOVA. While these analyses demonstrate differences between overview and diagnosis/treatment materials, they do not explore factors that may predict readability. Additional analyses evaluating the influence of institution type, document length, word count, or other characteristics could provide greater scientific insight. Multivariable modeling should be presented.

9.Choice of statistical test

Each participating institution contributed both an overview page and a diagnosis/treatment page, creating a naturally paired study design. However, the manuscript reports the use of an unpaired two-tailed t-test to compare these groups. The authors should justify why an unpaired analysis was selected rather than a paired t-test, which would generally be more appropriate for paired observations originating from the same institution. If there is a methodological reason for this choice, it should be clearly explained within the Methods section.

Overall, this is a well-written and technically competent descriptive study. Addressing the above concerns would improve the rigor of the manuscript, appropriately constrain the interpretation of the findings, and better position the work within the broader literature on online patient education.

Response to Reviewers

Manuscript Title: Readability of Online Melanoma Patient Education Materials

The author thanks the Editor and Reviewers for their thoughtful comments and constructive suggestions. All comments were carefully considered, and the manuscript was revised accordingly. Each reviewer comment is reproduced verbatim below, followed by a detailed response describing how and where the manuscript was revised.

Reviewer 1**Comment 1 – Limited novelty****Reviewer Comment**

The overall study design closely follows the methodology employed in numerous previous readability studies... The Discussion should better position the work within the existing literature and explicitly describe its incremental contribution.

Response

The author thanks the reviewer for this thoughtful and constructive comment. The author agrees that the present study represents an updated application of an established readability assessment methodology rather than the introduction of a novel analytical framework. In response to this comment, the author revised the Discussion to better position the study within the existing literature and to explicitly describe its incremental contribution. The Discussion was also expanded to describe the additional statistical analyses incorporated into the revised manuscript, including paired statistical comparisons, exploratory principal component analysis (PCA), and multivariable linear regression, thereby clarifying the study's incremental contribution.

The Discussion was revised to state:

"Although this study applied an established methodology for assessing readability, it provided an updated and incremental contribution to the existing literature by evaluating contemporary melanoma patient education materials from academic medical centers, professional organizations, and government sources. In addition, the incorporation of paired statistical comparisons, exploratory principal component analysis (PCA), and multivariable linear regression extended beyond a descriptive readability survey by quantitatively evaluating both the underlying structure of commonly used readability metrics and the independent factors associated with readability. PCA demonstrated that these readability indices largely measure a single underlying construct, supporting the consistency of commonly used readability formulas. Because the principal component analysis was exploratory and based on a relatively small sample of webpages, these findings should be interpreted cautiously and confirmed in larger studies evaluating a broader range of patient education materials. Taken together, these findings provided an updated assessment of the readability of sampled institutional online melanoma patient education materials while offering preliminary evidence regarding the interpretation and application of commonly used readability metrics."

These revisions were incorporated into the Discussion (pages 9–10). Corresponding methodological and statistical revisions were also incorporated into the Materials and Methods (pages 3–4) and Results (pages 7–9).

Comment 2 – Principal Component Analysis (PCA)**Reviewer Comment**

Given the conceptual overlap among readability formulas, the authors should consider performing principal component analysis (PCA) to better characterize the underlying structure of the readability metrics.

Response

The author thanks the reviewer for this excellent suggestion. In response to this comment, an exploratory principal component analysis (PCA) was performed using the five readability metrics included in the study. The Materials and Methods section was revised to describe the PCA methodology, including evaluation of sampling adequacy using the Kaiser–Meyer–Olkin (KMO) measure, Bartlett's test of sphericity, and parallel analysis to determine the number of retained

components. The Results section was expanded to report the PCA findings, and the Discussion was revised to interpret the significance of these results within the context of readability assessment. The Results section was revised to state:

"Exploratory principal component analysis retained a single principal component based on parallel analysis. The retained component had an eigenvalue of 4.71 and explained 94.2% of the total variance (Figure 1).

The overall Kaiser–Meyer–Olkin measure of sampling adequacy was 0.784, and Bartlett's test of sphericity demonstrated significant correlations among the readability metrics ($\chi^2(10) = 363$, $p < 0.001$). All grade-level readability indices loaded strongly and positively on the retained component (0.953–0.986), whereas Flesch Reading Ease loaded strongly in the opposite direction (–0.979), reflecting its inverse scoring system."

The Discussion was further revised to state:

"PCA demonstrated that these readability indices largely measure a single underlying construct, supporting the consistency of commonly used readability formulas. Because the principal component analysis was exploratory and based on a relatively small sample of webpages, these findings should be interpreted cautiously and confirmed in larger studies evaluating a broader range of patient education materials."

These revisions were incorporated into the Materials and Methods (pages 3–4), Results (pages 8–9), and Discussion (page 10).

Comment 3 – Limited representativeness of the sample

Reviewer Comment

The manuscript frequently generalizes its findings to online melanoma patient education materials as a whole. Because the sample was limited to selected institutional websites, the conclusions should be revised to clarify that the findings apply only to the evaluated sample and may not be generalizable to all online melanoma resources.

Response

The author thanks the reviewer for this important observation. In response to this comment, the Abstract, Materials and Methods, Discussion, and Conclusion were revised to consistently clarify that the findings apply only to the sampled institutional melanoma patient education materials rather than all online melanoma educational resources. Additional language was also added to acknowledge that the findings may not be generalizable to commercial websites, advocacy organizations, international organizations, or other educational resources that were not included in the study.

The Discussion was revised to state:

"This study has several limitations. First, it evaluated a sample of highly visible melanoma patient education materials from academic medical centers, professional organizations, and one government source. Accordingly, the findings should be interpreted as applying only to the sampled institutional resources and may not be generalizable to the broader online ecosystem, which includes commercial websites, advocacy organizations, international organizations, and other educational resources that were not evaluated."

The Conclusion was also revised to state:

"The sampled institutional melanoma patient education materials exceeded recommended readability levels and may be difficult for the average patient to understand."

These revisions were incorporated into the Abstract (page 1), Materials and Methods (pages 2–3), Discussion (pages 10–11), and Conclusion (page 11).

Comment 4 – Arbitrary website selection

Reviewer Comment

The rationale for selecting only the first 18 Google search results should be better justified.

Response

The author thanks the reviewer for this helpful suggestion. In response to this comment, the Materials and Methods section was substantially revised to better explain the sampling strategy. The revised manuscript now explains that academic medical centers were selected primarily from the

U.S. News & World Report Best Hospitals rankings and supplemented with professional organizations and a government health agency. The Materials and Methods section was further revised to clarify how replacement institutions were selected when eligible patient education materials were unavailable. Specifically, when an institution did not provide suitable publicly accessible melanoma patient education materials, it was replaced by the next highest-ranked academic medical center in the *U.S. News & World Report* Best Hospitals rankings that met the inclusion criteria. The assumptions underlying this sampling strategy were also explicitly described to clarify that the objective was to evaluate highly visible institutional melanoma patient education materials rather than exhaustively characterize all online melanoma educational resources.

The Materials and Methods section was revised to state:

"Academic medical centers were selected primarily from the U.S. News & World Report Best Hospitals rankings because these institutions are nationally recognized and highly visible to patients seeking online melanoma education (9). A total of 18 institutions, including academic medical centers, professional organizations, and government sources were selected due to their national ranking, providing enough data for the statistical analyses while maintaining a feasible scope for the analysis. When an institution did not provide suitable publicly accessible melanoma patient education materials, it was replaced by the next highest-ranked academic medical center in the U.S. News & World Report Best Hospitals rankings that met the inclusion criteria."

The Materials and Methods section was further revised to clarify how replacement institutions were selected when eligible patient education materials were unavailable. Specifically, institutions that did not provide suitable publicly accessible melanoma patient education materials were replaced by the next highest-ranked academic medical center in the U.S. News & World Report Best Hospitals rankings that met the inclusion criteria.

The Materials and Methods section further states:

"This sampling strategy assumed that nationally recognized institutional websites represent highly visible and commonly accessed sources of melanoma patient education. This assumption was considered reasonable because patients frequently rely on prominent academic medical centers, professional organizations, and government health agencies when seeking online health information."

These revisions were incorporated into the Materials and Methods (pages 2–3).

Comment 5 – Website rankings change over time

Reviewer Comment

Because Google search rankings change over time, the analyzed webpages represent only a snapshot of the online environment. This limitation should be acknowledged.

Response

The author thanks the reviewer for this important suggestion. In response to this comment, the Discussion was revised to acknowledge that Google search rankings are dynamic and may vary according to geographic location, personalization, browser history, and search algorithm updates. The manuscript now clarifies that the analyzed webpages represent a snapshot of the online environment at the time of data collection rather than a permanent ranking of melanoma educational resources.

The Discussion was revised to state:

"Furthermore, institution rankings are dynamic and may vary accordingly. Therefore, the analyzed PEMs represent a snapshot of the highest ranking at the time of data collection rather than a permanent ranking of available melanoma educational resources."

These revisions were incorporated into the Discussion (page 11).

Comment 6 – Readability is only one component of accessibility

Reviewer Comment

Readability is only one component of accessibility.

Response

The author thanks the reviewer for this thoughtful and constructive comment. The author agrees that readability represents only one component of accessibility and that readability formulas

estimate textual complexity rather than overall accessibility. In response to this comment, the Discussion was revised to distinguish readability from patient comprehension, and the Limitations section was expanded to acknowledge that additional factors influencing accessibility were beyond the scope of the present study.

The Discussion was revised to state:

"However, because readability formulas estimate textual complexity rather than patient comprehension, future studies involving patient testing are needed to determine whether improved readability leads to better patient understanding, decision-making, or clinical outcomes (3,5,10)."

The Limitations section was also revised to state:

"Additionally, readability formulas quantify textual complexity but do not directly measure patient comprehension or overall accessibility. Factors such as page organization, graphics, multimedia, cultural appropriateness, navigation, language availability, and content quality were not evaluated." These revisions were incorporated into the Discussion (page 10) and Limitations (page 11).

Comment 7 – No assessment of information quality

Reviewer Comment

The present study evaluates readability but does not assess the quality, completeness, or accuracy of the educational content itself. Established instruments such as DISCERN, PEMAT, and the JAMA benchmark criteria could provide complementary evaluations.

Response

The author thanks the reviewer for this valuable observation. The author agrees that readability alone does not evaluate the overall quality or usefulness of patient education materials. In response to this comment, the Limitations section was expanded to acknowledge that the present study did not evaluate the quality, completeness, understandability, or actionability of the educational content and to identify complementary assessment tools that may be incorporated into future research.

The Limitations section was revised to state:

"Similarly, this study did not assess the accuracy, completeness, understandability, or actionability of the educational content using instruments such as DISCERN, PEMAT, or the JAMA benchmark criteria."

These revisions were incorporated into the Limitations (page 11).

Comment 8 – No direct assessment of patient comprehension

Reviewer Comment

Although the manuscript appropriately notes that readability formulas do not directly measure patient comprehension, the conclusions should be tempered accordingly.

Response

The author thanks the reviewer for this insightful recommendation. The author agrees that readability should not be interpreted as a direct measure of patient comprehension or clinical outcomes. In response to this comment, the Discussion and Conclusion were revised to more clearly distinguish readability from patient comprehension and to emphasize that future studies involving patient testing are necessary to determine whether improved readability leads to improved patient understanding, decision-making, or clinical outcomes.

The Discussion was revised to state:

"However, because readability formulas estimate textual complexity rather than patient comprehension, future studies involving patient testing are needed to determine whether improved readability leads to better patient understanding, decision-making, or clinical outcomes (3,8,10)."

The Conclusion was revised to state:

"The sampled institutional melanoma patient education materials exceeded recommended readability levels and may be difficult for the average patient to understand. Given the importance of patient comprehension in melanoma care, these findings support continued efforts to develop patient education materials that are more accessible to a broad range of readers. As the availability of online health information continues to expand, improving the readability of melanoma PEMs may help support patient understanding and engagement. However, because readability is only one component of accessibility and does not directly measure patient comprehension, future studies

should evaluate whether improving readability leads to better patient understanding, decision-making, and other patient-centered outcomes through direct patient testing. Additional research should also examine the accessibility of melanoma patient education materials across multiple formats and diverse populations, including non-English resources.”

These revisions were incorporated into the Discussion (page 10) and Conclusion (page 12).

Comment 9 – Limited statistical analysis

Reviewer Comment

The statistical analyses are restricted primarily to descriptive statistics, t-tests, and ANOVA. Additional analyses evaluating factors that may predict readability would provide greater scientific insight.

Response

The author thanks the reviewer for this excellent suggestion. In response to this comment, the statistical analyses were expanded to include both an exploratory principal component analysis (PCA) and a multivariable linear regression analysis. The multivariable regression evaluated page category, institution type, and word count as independent predictors of readability. These additional analyses strengthened the statistical rigor of the study while providing greater insight into the factors independently associated with readability.

The Materials and Methods section was revised to state:

"A multivariable linear regression analysis was performed to identify factors independently associated with readability. The dependent variable was the average readability grade level, calculated as the mean of the Flesch–Kincaid Grade Level, SMOG Index, Gunning Fog Index, and Coleman–Liau Index. Independent variables included page category (Overview vs. Diagnosis & Treatment), institution type, and word count."

The Results section was revised to state:

"To further evaluate factors independently associated with readability, a multivariable linear regression analysis was performed using average readability grade level as the dependent variable. The regression model included page category (Overview vs. Diagnosis & Treatment), institution type, and word count as independent variables. Institution type was modeled as a categorical variable, with academic medical centers serving as the reference category. The overall regression model was statistically significant ($R^2 = 0.527$; adjusted $R^2 = 0.466$; $F(4,31) = 8.63$, $p < 0.001$). After adjusting for institution type and word count, page category remained the only significant independent predictor of readability."

The Discussion was also revised to state:

"Consistent with these findings, multivariable regression analysis demonstrated that page category remained independently associated with readability after adjustment for institution type and word count, whereas institution type and word count were not significant predictors."

These revisions were incorporated into the Materials and Methods (page 4), Results (pages 9–10), Discussion (page 11), and Table 4 (page 10).

Comment 10 – Choice of statistical test

Reviewer Comment

Each participating institution contributed both an overview page and a diagnosis/treatment page, creating a naturally paired study design. The authors should justify why an unpaired analysis was selected rather than a paired t-test.

Response

The author thanks the reviewer for identifying this issue. The author agrees that a paired statistical analysis is more appropriate because each participating institution contributed one overview page and one diagnosis/treatment page, creating naturally paired observations. Accordingly, the statistical analysis was revised by replacing the original independent-samples comparison with paired two-tailed t-tests throughout the manuscript, and the Materials and Methods, Results, and Table 3 were revised accordingly.

The Materials and Methods section was revised to state:

"To compare readability between melanoma overview and diagnosis/treatment PEMs, a paired two-tailed t-test was performed for each readability metric because each participating institution contributed one overview page and one diagnosis/treatment page, creating naturally paired observations."

The Results section was revised to state:

"Mean readability scores were significantly higher for diagnosis/treatment materials than overview materials across all readability metrics, as demonstrated by paired two-tailed t-tests (Table 3)."

These revisions were incorporated into the Materials and Methods (page 4), Results (pages 7–8), and Table 3 (pages 7–8).

Reviewer 2

Comment 1 – Verify that all links to the references point to the correct source.

Reviewer Comment

Please verify that all links to the references point to the correct source.

Response

The author thanks the reviewer for this careful review. In response to this comment, all references and hyperlinks were reviewed and verified to ensure that each citation directs readers to the appropriate source. Reference formatting was also reviewed for consistency and accuracy throughout the manuscript.

These revisions were incorporated into the References section (pages 12–13).

Comment 2 – Acknowledgments

Reviewer Comment

The authors must disclose and acknowledge any assistance received in the preparation of this manuscript, including but not limited to editorial, technical, analytical, or writing support. All such contributions must be clearly stated in the Acknowledgments section.

Response

The author thanks the reviewer for this important suggestion. In response to this comment, the Acknowledgements section was expanded to explicitly acknowledge the editorial feedback and writing support received during manuscript preparation while clarifying that the author was solely responsible for the study design, data collection, statistical analyses, interpretation of the findings, and preparation of the final manuscript.

The Acknowledgments section was revised to state:

"The author gratefully acknowledges and appreciates the editorial feedback and writing support provided by Dr. Austin Johnson and Shae Gantt during the preparation of this manuscript. The author was solely responsible for the study design, data collection, statistical analyses, interpretation of the findings, and preparation of the final manuscript."

These revisions were incorporated into the Acknowledgments section (page 12).

Comment 3 – References

Reviewer Comment

Include enough recent references along with foundational ones. Ensure references directly support your claims. Avoid "padding." Use credible sources (peer-reviewed journals, reputable books, official reports).

Response

The author thanks the reviewer for this valuable recommendation. In response to this comment, the reference list was carefully reviewed and revised to ensure that recent peer-reviewed literature was included alongside foundational references. References were evaluated to ensure that they directly supported the statements made throughout the manuscript, while unnecessary citations were avoided.

These revisions were incorporated throughout the Introduction (pages 1–2), Discussion (pages 10–11), Conclusion (pages 11–12), and References (pages 12–13).

Comment 4 – Assumptions

Reviewer Comment

All assumptions (including implicit assumptions) must be explicitly stated and clearly justified. The authors should explain why each assumption is reasonable and discuss its impact on the results and conclusions.

Response

The author thanks the reviewer for this thoughtful suggestion. In response to this comment, the Materials and Methods section was expanded to explicitly describe the assumptions underlying the study design, explain why these assumptions were considered reasonable, and clarify how they influence the interpretation of the findings.

The Materials and Methods section was revised to state:

"This sampling strategy assumed that nationally recognized institutional websites represent highly visible and commonly accessed sources of melanoma patient education. This assumption was considered reasonable because patients frequently rely on prominent academic medical centers, professional organizations, and government health agencies when seeking online health information. The study also assumed that established readability formulas provide valid estimates of textual complexity and that the selected webpages are representative of the educational content available from the participating institutions at the time of data collection."

The Materials and Methods section further states:

"Although these assumptions do not fully capture the diversity of online melanoma resources or all aspects of patient comprehension, they support a standardized comparison of widely accessible institutional materials. Consequently, the findings should be interpreted as applying to the sampled resources rather than the broader online landscape."

These revisions were incorporated into the Materials and Methods (pages 2–3).

Comment 5 – Overreaching conclusions**Reviewer Comment**

Avoid overreaching conclusions that extend beyond what is supported by the data and analysis.

Response

The author thanks the reviewer for this important observation. In response to this comment, the Discussion and Conclusion were revised to ensure that the conclusions remain appropriately limited to the findings of the present study. Statements implying direct effects on patient comprehension or clinical outcomes were tempered, and the manuscript now emphasizes that readability formulas estimate textual complexity rather than patient comprehension. The revised manuscript also highlights the need for future studies involving patient testing before conclusions regarding patient outcomes can be drawn.

The Discussion was revised to state:

"However, because readability formulas estimate textual complexity rather than patient comprehension, future studies involving patient testing are needed to determine whether improved readability leads to better patient understanding, decision-making, or clinical outcomes (3,8,10)."

The Conclusion was also revised to state:

"The sampled institutional melanoma patient education materials exceeded recommended readability levels and may be difficult for the average patient to understand... However, because readability is only one component of accessibility and does not directly measure patient comprehension, future studies should evaluate whether improving readability leads to better patient understanding, decision-making, and other patient-centered outcomes through direct patient testing. Additional research should also examine the accessibility of melanoma patient education materials across multiple formats and diverse populations, including non-English resources."

These revisions were incorporated into the Discussion (page 10) and Conclusion (page 11).

Comment 6 – Tense and third person**Reviewer Comment**

Verify that you have used past perfect tense and third person throughout the manuscript wherever applicable.

Response

The author thanks the reviewer for this careful review. In response to this comment, the manuscript was thoroughly reviewed for consistency of tense and point of view. Revisions were made throughout the manuscript to ensure that third-person language was used consistently and that verb tense was appropriate for each section. Descriptions of the study design, methodology, analyses, and findings were revised where appropriate to maintain consistent past tense, while present tense was retained where appropriate for established scientific knowledge and general statements. These revisions were incorporated throughout the Abstract (page 1), Introduction (pages 1–2), Materials and Methods (pages 2–4), Results (pages 4–9), Discussion (pages 9–11), and Conclusion (pages 11–12).

The author is commended for the substantial effort made in responding to the previous review. In particular, the addition of principal component analysis (PCA) represents a meaningful response to an important component of my earlier comments. The revised analysis now demonstrates that the five conventional readability measures are highly redundant: a single principal component explains 94.2% of the variance, with very strong loadings across the individual readability indices. This is useful because it establishes empirically that, within this dataset, the five readability formulas are predominantly measuring a common underlying dimension rather than five substantially independent properties of the text. I appreciate that the author undertook this additional analysis rather than simply discussing PCA as a future possibility.

However, the PCA result now creates a particularly good opportunity to take the analysis one important step further and make this paper a more credible and distinctive contribution to the literature. The purpose of my earlier PCA recommendation was not solely to determine whether the readability indices were correlated or reducible to fewer dimensions. I had also suggested projecting the individual webpages into principal-component space, identifying webpages that perform unusually well despite presenting complex clinical material, and examining the characteristics of those webpages. The present revision has completed the dimensional-reduction portion of that proposal, but not this subsequent and scientifically more informative portion.

Indeed, the result obtained by the author makes this extended analysis especially attractive. The PCA has shown that there are not several independent readability dimensions represented by these formulas; rather, there is essentially one dominant readability dimension. Therefore, the next scientific question becomes quite straightforward: what distinguishes webpages located at the favorable end of this empirically derived readability dimension from those at the unfavorable end? At present, the manuscript demonstrates that diagnosis/treatment pages are more difficult to read and that several readability formulas agree about this difference, but it does not determine what characteristics of the writing are responsible for the difference.

I therefore strongly encourage the author to calculate and report the PC1 score for each webpage and use these scores to characterize the distribution of individual resources along the dominant readability dimension. A figure displaying individual webpages according to their PC1 scores would be particularly informative. Webpages could be labeled or distinguished according to whether they are overview or diagnosis/treatment materials. This would immediately reveal whether particular resources perform substantially better or worse than others and whether there are potentially informative outliers.

The most interesting comparison would not necessarily be between the easiest overview page and the most difficult treatment page, because these resources contain inherently different types of information. Rather, the authors should identify high- and low-performing webpages within the same content category, particularly among diagnosis/treatment resources. The existing data already suggest that substantial differences exist within this category. For example, diagnosis/treatment FKGL values range from approximately 8.0 for NCI and 9.3 for Dana-Farber to 13.8 for UChicago and 14.1 for Johns Hopkins. Thus, even among webpages attempting to communicate relatively complex diagnosis and treatment information, some appear to accomplish this at a substantially lower estimated reading level than others. These contrasts potentially contain considerably more

scientific information than the overall observation that treatment pages are more difficult than overview pages.

The authors should therefore examine the textual characteristics associated with position on PC1. This analysis should move as far as reasonably possible toward the primitive characteristics of the writing itself, rather than simply recalculating additional composite readability scores. Candidate variables could include mean sentence length, median sentence length, variability in sentence length, proportion of very long sentences, mean word length, syllables per word, proportion of multisyllabic words, frequency of uncommon vocabulary, lexical diversity, density of medical terminology, frequency of abbreviations or acronyms, frequency of unexplained technical terms, passive-voice constructions, nominalizations, subordinate clauses or another measure of syntactic complexity, and paragraph length. The precise variables need not be limited to this list, and the author should select features that can be reproducibly quantified with the available text.

This is important because the existing readability formulas themselves are mathematical combinations of overlapping textual characteristics. Consequently, simply showing that FKGL, SMOG, Gunning Fog, Coleman–Liau, and Flesch Reading Ease correlate strongly is informative about redundancy among the measurement instruments but does not tell the reader what properties of the actual writing produce poor readability. The extended analysis should attempt to bridge that gap.

One possible approach would be to correlate each candidate textual feature with PC1. Given the relatively small number of webpages, the author should avoid fitting an excessively large multivariable model containing many correlated predictors. Instead, a sensible exploratory strategy would be to first quantify a predefined and reasonably limited set of linguistic features, examine their distributions and correlations with PC1, and then determine which features most strongly distinguish favorable from unfavorable readability. Confidence intervals and effect sizes would be preferable to relying exclusively on multiple hypothesis tests.

An even stronger approach, if the number of extracted linguistic variables permits it, would be to perform a second exploratory PCA or related dimensional-reduction analysis on the underlying textual features themselves. This analysis would answer a fundamentally different question from the PCA already performed. The existing PCA asks, “How many independent dimensions are represented by the five readability formulas?” A feature-level PCA would ask, “What independent dimensions of writing characterize these webpages?” For example, the data might reveal a dimension dominated by sentence/syntactic complexity, another dominated by vocabulary or medical terminology, and perhaps another related to structural composition. Whether these particular dimensions emerge should, of course, be determined by the data rather than assumed beforehand.

The authors could then determine which of these feature-derived dimensions are associated with the existing readability PC1 score and which distinguish overview from diagnosis/treatment pages. This would provide a much more informative conceptual framework. Instead of merely concluding that diagnosis/treatment pages receive worse scores on several highly correlated formulas, the study could begin to establish whether their increased difficulty is predominantly associated with sentence structure, lexical complexity, technical terminology, grammatical construction, or combinations of these characteristics.

This analysis would also provide an opportunity to investigate positive outliers, which was an important aspect of my earlier recommendation. A diagnosis/treatment webpage that communicates complex clinical information but nevertheless has a favorable PC1 score is scientifically valuable because it may represent a naturally occurring example of effective patient communication. The authors should identify such webpages and ask what differentiates them from diagnosis/treatment webpages at the unfavorable end of the distribution. Conversely, an unusually difficult overview page could also be informative.

This need not become a subjective assessment of which institution “writes better.” The analysis should remain quantitative. For example, if favorable diagnosis/treatment webpages contain shorter sentences, fewer multisyllabic words, fewer unexplained medical terms, less syntactic nesting, or a

different balance of technical and ordinary vocabulary, these characteristics can be measured and compared. The objective is not to rank institutions but to identify potentially transferable writing strategies.

This would also strengthen an important weakness in the present Discussion. The manuscript currently suggests that the greater difficulty of diagnosis/treatment materials is “likely reflecting the inherent complexity of diagnosis and treatment information.” This explanation is plausible, but it is not directly demonstrated by the present analysis. The proposed extended analysis could begin to test this explanation. If treatment pages contain substantially more technical terminology but comparable sentence structures, for example, that would suggest one mechanism. If terminology density is similar but treatment pages contain longer and syntactically more complex sentences, that would suggest another. If both differ, their relative contributions could be explored.

The same issue applies to the recommendations currently made in the Discussion. The manuscript recommends simplifying medical terminology, reducing sentence length, and minimizing complex multisyllabic words. These are reasonable recommendations, but at present they are largely general readability principles rather than findings generated by this particular study. The proposed analysis would allow the authors to determine whether these characteristics actually distinguish the more readable melanoma PEMs in their dataset. The recommendations could then become empirically supported conclusions of the study rather than generic recommendations appended to the findings. I would also encourage the authors to consider whether some high-performing webpages achieve favorable readability without simply eliminating medically necessary information. This is potentially one of the most interesting questions raised by the dataset. The ideal patient education resource is not merely one with short words and short sentences; it must communicate clinically important information accurately while remaining understandable. If a diagnosis/treatment page achieves a favorable PC1 score while still discussing complex topics, examining how it accomplishes this could provide an evidence-based model for future PEM development. The present study did not assess completeness or accuracy using DISCERN, PEMAT, or JAMA criteria, and the manuscript appropriately acknowledges this limitation. Therefore, the authors should not claim that a lower-readability-score page necessarily provides superior education. Nevertheless, they can legitimately investigate the linguistic strategies associated with reduced textual complexity while clearly maintaining this limitation.

The revised analysis should therefore distinguish between readability performance and overall quality of the educational resource. A favorable PC1 score means that a webpage performs favorably along the common readability dimension identified by these formulas; it does not establish accuracy, completeness, actionability, or actual patient comprehension. Maintaining this distinction would preserve the appropriately cautious interpretation that the authors have now incorporated elsewhere in the manuscript.

Because there are only 36 webpages, I do not expect an excessively elaborate natural-language-processing study or a highly parameterized machine-learning model. In fact, such approaches could easily overfit this relatively small dataset. A carefully designed exploratory analysis using a modest number of transparent, reproducible textual features would be scientifically preferable. The purpose should be to extract additional information from the existing dataset rather than to make strong predictive claims from a small sample.

The authors should also avoid defining “high-performing” and “low-performing” webpages solely through arbitrary cutoffs if possible. PC1 can initially be treated as a continuous outcome. Feature-PC1 relationships can then be examined across all webpages, followed by illustrative examination of resources at the extremes of the distribution. If the authors choose to compare groups—for example, the most favorable and least favorable quartiles—this should be presented as an exploratory analysis and the threshold should be defined in advance and clearly reported.

The present PCA therefore should not be removed or replaced. It provides the logical foundation for the proposed extension. The analytical sequence could become:

individual readability indices → PCA of readability indices → dominant readability PC1 → individual webpage PC1 scores → quantitative textual features → identification of textual

characteristics associated with favorable PC1 → examination of high-performing diagnosis/treatment resources → evidence-based recommendations for future melanoma PEM design.

This would provide a much stronger narrative than treating the PCA as an endpoint.

The revised manuscript has already established an important empirical result: the five conventional readability measures collapse overwhelmingly onto one common dimension. The authors now have an opportunity to capitalize on that finding. Instead of concluding simply that multiple formulas consistently classify these materials as difficult, the paper could determine what writing characteristics move a webpage toward the favorable or unfavorable end of that common readability dimension.

This would substantially improve the novelty and scientific value of the manuscript. The central question would evolve from the already well-established observation that online medical education materials are frequently written above recommended reading levels to the considerably more useful question:

What characteristics distinguish melanoma patient education materials that communicate complex clinical information with comparatively favorable readability, and can those characteristics provide an empirical framework for designing better patient education materials?

That is, in my view, the most promising direction available from the present dataset. The author has made substantial progress in the revision and should be commended for undertaking the PCA and the other requested statistical improvements. Rather than viewing this additional request as a rejection of the work already performed, I encourage the author to regard it as an opportunity created by the very strong PCA result. The analysis has now reached the point where a relatively focused extension could move the manuscript from another confirmatory readability survey toward a study that begins to explain why some patient education materials are more readable than others and how that information might be used prospectively to improve their design. I therefore recommend revision, with particular emphasis on extending the PCA analysis in this direction.

Response to Reviewers

Manuscript Title: Readability of Online Melanoma Patient Education Materials

The author thanks the Editor and Reviewer for the thoughtful comments and constructive suggestions. All comments were carefully considered, and the manuscript was revised accordingly. Each reviewer comment is reproduced verbatim below, followed by a detailed response describing how and where the manuscript was revised.

Reviewer 1 – Second-Round Review

Comment 1 – Commendation regarding the prior revision and PCA

Reviewer Comment

The author is commended for the substantial effort made in responding to the previous review. In particular, the addition of principal component analysis (PCA) represents a meaningful response to an important component of my earlier comments. The revised analysis now demonstrates that the five conventional readability measures are highly redundant: a single principal component explains 94.2% of the variance, with very strong loadings across the individual readability indices. This is useful because it establishes empirically that, within this dataset, the five readability formulas are predominantly measuring a common underlying dimension rather than five substantially independent properties of the text. I appreciate that the author undertook this additional analysis rather than simply discussing PCA as a future possibility.

Response

The author sincerely thanks the reviewer for this encouraging assessment. The original readability-index PCA was retained as the analytical foundation for the extended revision. During final reproducibility checks in Jamovi, the PCA was rerun and the corresponding values were updated slightly. The retained component had an eigenvalue of 4.72 and explained 94.3% of the total

variance, with an overall KMO of 0.788. The grade-level indices loaded strongly and positively (0.954–0.986), whereas Flesch Reading Ease loaded negatively (–0.979). These updated values are now reported consistently in the Results, without changing the substantive interpretation that the five formulas largely reflect one common readability construct.

Results, pages 9–10; Figure 1, page 9.

Comment 2 – Extend the PCA beyond dimensional reduction

Reviewer Comment

However, the PCA result now creates a particularly good opportunity to take the analysis one important step further and make this paper a more credible and distinctive contribution to the literature. The purpose of my earlier PCA recommendation was not solely to determine whether the readability indices were correlated or reducible to fewer dimensions. I had also suggested projecting the individual webpages into principal-component space, identifying webpages that perform unusually well despite presenting complex clinical material, and examining the characteristics of those webpages. The present revision has completed the dimensional-reduction portion of that proposal, but not this subsequent and scientifically more informative portion.

Response

The author agrees and substantially extended the analysis in this direction. Individual PC1 scores were calculated for all 36 webpages and plotted along the dominant readability dimension. The revised analysis then quantified reproducible sentence- and word-level features, examined their continuous associations with PC1, compared those features between paired overview and diagnosis/treatment pages, descriptively examined diagnosis/treatment webpages at the favorable and unfavorable ends of PC1, and performed a second exploratory PCA on the quantified textual features. The Abstract, Materials and Methods, Results, Discussion, Limitations, recommendations, and Conclusion were revised to incorporate this extension.

Abstract, page 1; Materials and Methods, pages 4–5; Results, pages 11–14; Discussion and Limitations, pages 15–16; Conclusion, page 17.

Comment 3 – Identify what distinguishes favorable from unfavorable PC1 performance

Reviewer Comment

Indeed, the result obtained by the author makes this extended analysis especially attractive. The PCA has shown that there are not several independent readability dimensions represented by these formulas; rather, there is essentially one dominant readability dimension. Therefore, the next scientific question becomes quite straightforward: what distinguishes webpages located at the favorable end of this empirically derived readability dimension from those at the unfavorable end? At present, the manuscript demonstrates that diagnosis/treatment pages are more difficult to read and that several readability formulas agree about this difference, but it does not determine what characteristics of the writing are responsible for the difference.

Response

The revised manuscript directly addresses this question by linking PC1 to measurable textual characteristics. Greater textual difficulty was most strongly associated with a higher proportion of multisyllabic words, greater mean word length, a greater proportion of sentences exceeding 20 words, and greater mean sentence length. Paired comparisons further showed that diagnosis/treatment materials had significantly longer sentences, more sentences exceeding 20 words, longer words, and a higher proportion of multisyllabic words than their paired overview materials. The Discussion was revised so that the observed readability difference is interpreted in terms of these measurable writing characteristics rather than being attributed only to the inherent complexity of clinical content.

Results, pages 12–14 (Table 5 and paired comparisons); Discussion, page 15.

Comment 4 – Calculate and report individual webpage PC1 scores and add a figure

Reviewer Comment

I therefore strongly encourage the author to calculate and report the PC1 score for each webpage and use these scores to characterize the distribution of individual resources along the dominant readability dimension. A figure displaying individual webpages according to their PC1 scores would be particularly informative. Webpages could be labeled or distinguished according to whether they are overview or diagnosis/treatment materials. This would immediately reveal whether particular resources perform substantially better or worse than others and whether there are potentially informative outliers.

Response

This recommendation was implemented. Individual PC1 scores were calculated for all 36 webpages, with PC1 oriented so that higher values indicated greater textual difficulty and lower values indicated more favorable readability. Figure 2 displays all 36 webpages along the PC1 continuum, with circles representing overview materials and triangles representing diagnosis/treatment materials. The Results also report exact PC1 values for representative diagnosis/treatment resources at the favorable and unfavorable ends of the distribution. The Figure 2 caption explicitly cautions that PC1 represents readability performance rather than educational quality, accuracy, completeness, actionability, or patient comprehension.

Materials and Methods, page 4; Results, pages 11–12; Figure 2, page 12.

Comment 5 – Compare webpages within the same content category

Reviewer Comment

The most interesting comparison would not necessarily be between the easiest overview page and the most difficult treatment page, because these resources contain inherently different types of information. Rather, the authors should identify high- and low-performing webpages within the same content category, particularly among diagnosis/treatment resources. The existing data already suggest that substantial differences exist within this category. For example, diagnosis/treatment FKGL values range from approximately 8.0 for NCI and 9.3 for Dana-Farber to 13.8 for UChicago and 14.1 for Johns Hopkins. Thus, even among webpages attempting to communicate relatively complex diagnosis and treatment information, some appear to accomplish this at a substantially lower estimated reading level than others. These contrasts potentially contain considerably more scientific information than the overall observation that treatment pages are more difficult than overview pages.

Response

The revised analysis focuses the illustrative extreme comparison within the diagnosis/treatment category, as recommended. NCI (PC1 = -0.979), Dana-Farber (PC1 = -0.625), and UPMC (PC1 = -0.604) were identified at the more favorable end of the diagnosis/treatment PC1 distribution, whereas UChicago (PC1 = 1.487), Brown Health (PC1 = 1.578), and Johns Hopkins (PC1 = 2.202) were identified at the less favorable end. Their sentence- and word-level characteristics were then compared descriptively. These examples were explicitly presented as illustrative contrasts rather than institutional rankings.

Results, pages 11 and 14.

Comment 6 – Quantify primitive textual characteristics

Reviewer Comment

The authors should therefore examine the textual characteristics associated with position on PC1. This analysis should move as far as reasonably possible toward the primitive characteristics of the writing itself, rather than simply recalculating additional composite readability scores. Candidate variables could include mean sentence length, median sentence length, variability in sentence length, proportion of very long sentences, mean word length, syllables per word, proportion of multisyllabic words, frequency of uncommon vocabulary, lexical diversity, density of medical terminology, frequency of abbreviations or acronyms, frequency of unexplained technical terms,

passive-voice constructions, nominalizations, subordinate clauses or another measure of syntactic complexity, and paragraph length. The precise variables need not be limited to this list, and the author should select features that can be reproducibly quantified with the available text.

Response

The author implemented a predefined, limited set of transparent and reproducibly quantifiable textual features appropriate to the sample size and available cleaned text. These included mean sentence length, the percentage of sentences containing more than 20 words, mean word length, the percentage of words containing three or more syllables, moving-average type-token ratio using a 50-word window (MATTR-50) as a measure of lexical diversity, and acronym frequency per 1,000 words. Operational definitions were added to the Materials and Methods.

The selected feature set was intentionally modest and reproducible, consistent with the reviewer's later caution against overparameterization in this 36-webpage dataset.

Materials and Methods, pages 4–5.

Comment 7 – Bridge readability-formula redundancy to properties of the writing

Reviewer Comment

This is important because the existing readability formulas themselves are mathematical combinations of overlapping textual characteristics. Consequently, simply showing that FKGL, SMOG, Gunning Fog, Coleman–Liau, and Flesch Reading Ease correlate strongly is informative about redundancy among the measurement instruments but does not tell the reader what properties of the actual writing produce poor readability. The extended analysis should attempt to bridge that gap.

Response

The revised analysis was designed specifically to bridge this gap. Rather than adding additional composite readability formulas, the manuscript relates the dominant readability PC1 to directly quantified sentence- and word-level features. Table 5 reports continuous Pearson correlations between each textual feature and PC1 with 95% confidence intervals, while the paired feature analysis identifies which writing characteristics differ between overview and diagnosis/treatment materials. The Discussion now interprets the findings in terms of sentence-level and lexical complexity.

Materials and Methods, page 4; Results, pages 12–14 (Table 5); Discussion, page 15.

Comment 8 – Use continuous feature-PC1 relationships with confidence intervals and effect sizes

Reviewer Comment

One possible approach would be to correlate each candidate textual feature with PC1. Given the relatively small number of webpages, the author should avoid fitting an excessively large multivariable model containing many correlated predictors. Instead, a sensible exploratory strategy would be to first quantify a predefined and reasonably limited set of linguistic features, examine their distributions and correlations with PC1, and then determine which features most strongly distinguish favorable from unfavorable readability. Confidence intervals and effect sizes would be preferable to relying exclusively on multiple hypothesis tests.

Response

The reviewer's proposed strategy was followed. PC1 was treated as a continuous variable and Pearson correlations with 95% confidence intervals were calculated for each of six predefined textual features. The strongest associations with greater textual difficulty were multisyllabic-word percentage ($r = 0.900$, 95% CI 0.812–0.948), mean word length ($r = 0.871$, 95% CI 0.760–0.933), long-sentence percentage ($r = 0.785$, 95% CI 0.615–0.885), and mean sentence length ($r = 0.735$, 95% CI 0.535–0.856). MATTR-50 showed a moderate association ($r = 0.611$), whereas acronym frequency was not significantly associated with PC1 ($r = 0.060$). For paired overview-versus-diagnosis/treatment feature comparisons, 95% confidence intervals and Cohen's d_z effect sizes were reported for the significant differences. No large multivariable linguistic model was fitted. The Materials and Methods section was also revised to specify that paired two-tailed t -

tests were used for the feature comparisons and that 95% confidence intervals and Cohen's *d* were reported.

Materials and Methods, page 4; Results, pages 12–14; Table 5, page 13.

Comment 9 – Perform an exploratory feature-level PCA

Reviewer Comment

An even stronger approach, if the number of extracted linguistic variables permits it, would be to perform a second exploratory PCA or related dimensional-reduction analysis on the underlying textual features themselves. This analysis would answer a fundamentally different question from the PCA already performed. The existing PCA asks, “How many independent dimensions are represented by the five readability formulas?” A feature-level PCA would ask, “What independent dimensions of writing characterize these webpages?” For example, the data might reveal a dimension dominated by sentence/syntactic complexity, another dominated by vocabulary or medical terminology, and perhaps another related to structural composition. Whether these particular dimensions emerge should, of course, be determined by the data rather than assumed beforehand.

Response

The author performed the suggested second exploratory PCA on the quantified textual features. The initial analysis included all six features. Acronym frequency demonstrated poor individual sampling adequacy and minimal association with the other textual features, so this diagnostic finding was reported transparently and acronym frequency was excluded from the final feature-level PCA while remaining in the correlation and paired analyses. Parallel analysis retained one component among the remaining five features (eigenvalue = 3.46), explaining 69.1% of the variance. The overall KMO was 0.687 and Bartlett's test was significant ($\chi^2(10) = 192, p < 0.001$). All five retained features loaded positively, indicating that the measured writing characteristics primarily reflected one broad linguistic-complexity dimension rather than several independent dimensions. Component scores from the retained feature-level PCA were then used in subsequent analyses to determine whether this linguistic-complexity dimension was associated with the original readability PC1 and whether it distinguished overview from diagnosis/treatment materials.

Materials and Methods, page 5; Results, page 14; Discussion, page 15.

Comment 10 – Relate feature-derived dimensions to readability and page category

Reviewer Comment

The authors could then determine which of these feature-derived dimensions are associated with the existing readability PC1 score and which distinguish overview from diagnosis/treatment pages. This would provide a much more informative conceptual framework. Instead of merely concluding that diagnosis/treatment pages receive worse scores on several highly correlated formulas, the study could begin to establish whether their increased difficulty is predominantly associated with sentence structure, lexical complexity, technical terminology, grammatical construction, or combinations of these characteristics.

Response

The author thanks the reviewer for this important suggestion. The revised analysis specifically evaluated sentence-level and lexical characteristics to determine which aspects of writing were associated with greater readability difficulty. Sentence-level complexity was represented by mean sentence length and the proportion of sentences exceeding 20 words, while lexical complexity was represented by mean word length, the proportion of multisyllabic words, and lexical diversity measured by MATTR-50. These features were all examined in relation to the original readability PC1. The strongest associations with greater readability difficulty were observed for multisyllabic-word percentage, mean word length, long-sentence percentage, and mean sentence length. The feature-level PCA further showed that these measured sentence-level and lexical characteristics did not separate into independent dimensions. Instead, they converged on a single general linguistic-complexity component, with positive loadings for multisyllabic-word percentage (0.904), long-

sentence percentage (0.871), mean word length (0.869), mean sentence length (0.812), and MATTR-50 (0.683). This finding suggests that increased readability difficulty in the sampled webpages was associated with a combination of sentence-level and lexical complexity rather than a single isolated characteristic. Technical terminology density and grammatical-construction measures were not directly quantified in the present analysis because the feature set was intentionally limited to transparent, reproducible measures appropriate to the relatively small sample size.

The feature-derived linguistic-complexity component was also strongly associated with the original readability PC1 ($r = 0.946$, 95% CI: 0.895–0.972; $p < 0.001$). In paired comparisons, diagnosis/treatment materials had significantly higher linguistic-complexity component scores than overview materials (0.507 vs. -0.507; mean difference = 1.014, 95% CI: 0.532–1.496; $t(17) = 4.40$, $p < 0.001$; Cohen's $d_z = 1.05$). Together, these results indicate that the increased difficulty of diagnosis/treatment materials was associated primarily with a combined pattern of greater sentence and lexical complexity among the features evaluated.

Materials and Methods, page 5; Results, page 14; Discussion, page 15.

Comment 11 – Investigate favorable diagnosis/treatment outliers

Reviewer Comment

This analysis would also provide an opportunity to investigate positive outliers, which was an important aspect of my earlier recommendation. A diagnosis/treatment webpage that communicates complex clinical information but nevertheless has a favorable PC1 score is scientifically valuable because it may represent a naturally occurring example of effective patient communication. The authors should identify such webpages and ask what differentiates them from diagnosis/treatment webpages at the unfavorable end of the distribution. Conversely, an unusually difficult overview page could also be informative.

Response

The revised Results identify diagnosis/treatment resources at both ends of the PC1 distribution and examine their textual characteristics descriptively. Favorable diagnosis/treatment examples (NCI, Dana-Farber, and UPMC) had shorter mean sentence lengths, fewer sentences exceeding 20 words, shorter mean word lengths, and lower multisyllabic-word percentages than the less favorable diagnosis/treatment examples (UChicago, Brown Health, and Johns Hopkins). The primary illustrative outlier analysis was intentionally focused on diagnosis/treatment materials because the reviewer identified this within-category comparison as the most scientifically informative. The analysis remained descriptive and did not assign formal high/low categories to all webpages.

Results, pages 11 and 14.

Comment 12 – Keep the comparison quantitative and avoid institutional ranking

Reviewer Comment

This need not become a subjective assessment of which institution “writes better.” The analysis should remain quantitative. For example, if favorable diagnosis/treatment webpages contain shorter sentences, fewer multisyllabic words, fewer unexplained medical terms, less syntactic nesting, or a different balance of technical and ordinary vocabulary, these characteristics can be measured and compared. The objective is not to rank institutions but to identify potentially transferable writing strategies.

Response

The manuscript was revised to keep the comparison quantitative and to avoid claims that one institution provides superior education. The extreme diagnosis/treatment examples are described using measured mean sentence length, long-sentence percentage, mean word length, and multisyllabic-word percentage. The Results explicitly state that these descriptive contrasts are illustrative examples rather than formal institutional rankings. The Limitations further state that favorable PC1 performance should not be interpreted as superior patient education.

Results, page 14; Limitations, page 16.

Comment 13 – Replace the speculative explanation of diagnosis/treatment difficulty

Reviewer Comment

This would also strengthen an important weakness in the present Discussion. The manuscript currently suggests that the greater difficulty of diagnosis/treatment materials is “likely reflecting the inherent complexity of diagnosis and treatment information.” This explanation is plausible, but it is not directly demonstrated by the present analysis. The proposed extended analysis could begin to test this explanation. If treatment pages contain substantially more technical terminology but comparable sentence structures, for example, that would suggest one mechanism. If terminology density is similar but treatment pages contain longer and syntactically more complex sentences, that would suggest another. If both differ, their relative contributions could be explored.

Response

The speculative Discussion statement in the prior manuscript was removed and replaced with a data-driven interpretation. The revised Discussion now states that greater PC1 difficulty was associated with measurable sentence- and word-level characteristics and that diagnosis/treatment materials showed significantly greater sentence and word complexity than their paired overview materials. The manuscript therefore no longer attributes the difference simply to the inherent complexity of diagnosis/treatment content; instead, it states that the greater difficulty was associated not only with the clinical content being communicated but also with measurable characteristics of how that information was written.

Discussion, page 15.

Comment 14 – Make the recommendations empirically supported

Reviewer Comment

The same issue applies to the recommendations currently made in the Discussion. The manuscript recommends simplifying medical terminology, reducing sentence length, and minimizing complex multisyllabic words. These are reasonable recommendations, but at present they are largely general readability principles rather than findings generated by this particular study. The proposed analysis would allow the authors to determine whether these characteristics actually distinguish the more readable melanoma PEMs in their dataset. The recommendations could then become empirically supported conclusions of the study rather than generic recommendations appended to the findings.

Response

The recommendations were revised so that they arise directly from the study’s observed associations. The revised Discussion states that more favorable readability was associated with shorter sentences, fewer sentences exceeding 20 words, shorter average word length, and a lower proportion of multisyllabic words. It then recommends shortening sentence structure where clinically appropriate and replacing unnecessarily complex terminology with simpler alternatives while preserving medically necessary information. The Conclusion likewise identifies sentence length and lexical complexity as potentially modifiable writing characteristics. *Discussion, page 16; Conclusion, page 17.*

Comment 15 – Consider favorable readability while preserving medically necessary information

Reviewer Comment

I would also encourage the authors to consider whether some high-performing webpages achieve favorable readability without simply eliminating medically necessary information. This is potentially one of the most interesting questions raised by the dataset. The ideal patient education resource is not merely one with short words and short sentences; it must communicate clinically important information accurately while remaining understandable. If a diagnosis/treatment page achieves a favorable PC1 score while still discussing complex topics, examining how it accomplishes this could provide an evidence-based model for future PEM development. The present study did not assess completeness or accuracy using DISCERN, PEMAT, or JAMA criteria, and the manuscript appropriately acknowledges this limitation.

Therefore, the authors should not claim that a lower-readability-score page necessarily provides superior education. Nevertheless, they can legitimately investigate the linguistic strategies associated with reduced textual complexity while clearly maintaining this limitation.

Response

The author agrees with this distinction. Because the study did not include a validated assessment of content accuracy, completeness, actionability, or understandability, the revised manuscript does not claim that favorable-PC1 resources preserve more medically necessary information or provide superior education. Instead, the Results use favorable and unfavorable diagnosis/treatment resources as descriptive examples to examine linguistic strategies associated with reduced textual complexity, including differences in sentence length, long-sentence frequency, mean word length, and multisyllabic-word frequency. The Discussion recommendations explicitly state that simplification should occur while preserving medically necessary information, and the Limitations clearly state that educational quality and patient comprehension were not assessed. This approach allows the linguistic comparison requested by the reviewer without exceeding what the data can support.

Results, page 14; Discussion and Limitations, page 16.

Comment 16 – Distinguish readability performance from educational quality

Reviewer Comment

The revised analysis should therefore distinguish between readability performance and overall quality of the educational resource. A favorable PC1 score means that a webpage performs favorably along the common readability dimension identified by these formulas; it does not establish accuracy, completeness, actionability, or actual patient comprehension. Maintaining this distinction would preserve the appropriately cautious interpretation that the authors have now incorporated elsewhere in the manuscript.

Response

This distinction was made explicit in both Figure 2 and the Limitations. The Figure 2 caption states that PC1 reflects readability performance and should not be interpreted as a measure of educational accuracy, completeness, actionability, quality, or patient comprehension. The Limitations further state that favorable PC1 performance is not evidence that a webpage provides superior patient education and that the identified resources should be interpreted as examples of comparatively favorable textual readability rather than rankings of educational quality.

Figure 2 caption, page 12; Limitations, page 16.

Comment 17 – Avoid overfitting and use a modest, transparent feature set

Reviewer Comment

Because there are only 36 webpages, I do not expect an excessively elaborate natural-language-processing study or a highly parameterized machine-learning model. In fact, such approaches could easily overfit this relatively small dataset. A carefully designed exploratory analysis using a modest number of transparent, reproducible textual features would be scientifically preferable. The purpose should be to extract additional information from the existing dataset rather than to make strong predictive claims from a small sample.

Response

The revised analysis follows this recommendation. Only six transparent and reproducibly defined linguistic features were extracted, and no machine-learning model or highly parameterized predictive model was fitted. The Materials and Methods explicitly state that the features were selected to capture sentence- and word-level complexity while limiting the number of variables given the relatively small sample size. The feature-level PCA is described as exploratory, and the Discussion continues to caution that the PCA findings should be confirmed in larger samples.

Materials and Methods, pages 4–5; Discussion, page 15.

Comment 18 – Treat PC1 continuously and avoid arbitrary high/low cutoffs

Reviewer Comment

The authors should also avoid defining “high-performing” and “low-performing” webpages solely through arbitrary cutoffs if possible. PC1 can initially be treated as a continuous outcome.

Feature-PC1 relationships can then be examined across all webpages, followed by illustrative examination of resources at the extremes of the distribution. If the authors choose to compare groups—for example, the most favorable and least favorable quartiles—this should be presented as an exploratory analysis and the threshold should be defined in advance and clearly reported.

Response

The author followed this recommendation directly. PC1 was analyzed as a continuous outcome across all 36 webpages, and feature-PC1 relationships were evaluated using continuous Pearson correlations with 95% confidence intervals. No arbitrary quartile, median, or other high/low threshold was used for inferential testing. Resources at the extremes of the diagnosis/treatment distribution were examined only descriptively as illustrative examples, and the manuscript explicitly states that these contrasts were not formal institutional rankings.

Materials and Methods, page 4; Results, pages 12–14.

Comment 19 – Retain the original readability PCA

Reviewer Comment

The present PCA therefore should not be removed or replaced. It provides the logical foundation for the proposed extension. The analytical sequence could become: **Response**

The original PCA of the five readability indices was retained and remains the logical foundation of the extended analysis. It appears in the Results before the individual PC1 projection and textual-feature analyses. The second feature-level PCA was added as a separate exploratory analysis rather than replacing the original readability PCA.

Materials and Methods, pages 4–5; Results, pages 9–14.

Comment 20 – Follow the proposed analytical sequence

Reviewer Comment

individual readability indices → PCA of readability indices → dominant readability PC1 → individual webpage PC1 scores → quantitative textual features → identification of textual characteristics associated with favorable PC1 → examination of high-performing diagnosis/treatment resources → evidence-based recommendations for future melanoma PEM design.

Response

The revised manuscript now follows this analytical sequence closely. It first reports the individual readability indices and paired comparisons (Tables 1–3), then the readability-index PCA and Figure 1, followed by the multivariable regression analysis (Table 4), individual webpage PC1 scores and Figure 2, quantified textual-feature correlations (Table 5), paired feature comparisons, descriptive examination of favorable and unfavorable diagnosis/treatment resources, the exploratory feature-level PCA, and finally data-supported recommendations in the Discussion and Conclusion.

Results, pages 5–14; Discussion, pages 15–16; Conclusion, page 17.

Comment 21 – Use the PCA as a foundation rather than an endpoint

Reviewer Comment

This would provide a much stronger narrative than treating the PCA as an endpoint.

Response

The author agrees. The revised manuscript no longer treats PCA as the analytical endpoint. The readability PCA now serves as the foundation for the individual webpage PC1 analysis, the textual-feature analysis, the within-category diagnosis/treatment comparison, and the

evidence-based recommendations that follow.
Results, pages 9–14; Discussion, pages 15–16.

Comment 22 – Use the common readability dimension to explain why materials differ
Reviewer Comment

The revised manuscript has already established an important empirical result: the five conventional readability measures collapse overwhelmingly onto one common dimension. The authors now have an opportunity to capitalize on that finding. Instead of concluding simply that multiple formulas consistently classify these materials as difficult, the paper could determine what writing characteristics move a webpage toward the favorable or unfavorable end of that common readability dimension.

Response

The revised manuscript capitalizes on the common readability dimension by identifying the textual characteristics associated with movement along PC1. The strongest correlates of greater difficulty were multisyllabic-word percentage, mean word length, long-sentence percentage, and mean sentence length. The paired analysis and descriptive diagnosis/treatment examples reinforce these findings. The Discussion and Conclusion were revised to emphasize these potentially modifiable writing characteristics rather than merely stating that several formulas agree that the materials are difficult.

Results, pages 12–14; Discussion and Conclusion, pages 15–17.

Comment 23 – Increase novelty and scientific value through explanatory analysis

Reviewer Comment

This would substantially improve the novelty and scientific value of the manuscript. The central question would evolve from the already well-established observation that online medical education materials are frequently written above recommended reading levels to the considerably more useful question:

Response

The author agrees and revised the manuscript's emphasis accordingly. The Abstract now states that the study examined textual characteristics associated with favorable and unfavorable readability performance, and the Results and Discussion move beyond confirming elevated reading levels to identify specific sentence- and word-level characteristics associated with the dominant readability dimension. The revised Conclusion likewise emphasizes potentially modifiable writing characteristics that may help guide future melanoma PEM development.

Abstract, page 1; Results, pages 12–14; Discussion and Conclusion, pages 15–17.

Comment 24 – Central scientific question proposed by the reviewer

Reviewer Comment

What characteristics distinguish melanoma patient education materials that communicate complex clinical information with comparatively favorable readability, and can those characteristics provide an empirical framework for designing better patient education materials? **Response**

The revised manuscript directly addresses this question within the limits of the dataset. The Introduction was revised to frame this question as a central motivation for the study by stating that the analysis sought to identify measurable writing characteristics associated with comparatively favorable readability and to explore whether those characteristics could provide an empirical framework for designing more readable melanoma patient education materials. In the Results, more favorable readability was associated with shorter sentences, fewer sentences exceeding 20 words, shorter average word length, and a lower proportion of multisyllabic words. These findings were translated into specific, empirically supported recommendations for reducing textual complexity while preserving medically necessary information. The manuscript remains cautious that readability alone does not establish educational quality or patient comprehension.

Introduction, page 2; Results, pages 12–14; Discussion and Limitations, pages 15–16; Conclusion, page 17.

Comment 25 – Overall recommendation to extend the PCA analysis

Reviewer Comment

That is, in my view, the most promising direction available from the present dataset. The author has made substantial progress in the revision and should be commended for undertaking the PCA and the other requested statistical improvements. Rather than viewing this additional request as a rejection of the work already performed, I encourage the author to regard it as an opportunity created by the very strong PCA result. The analysis has now reached the point where a relatively focused extension could move the manuscript from another confirmatory readability survey toward a study that begins to explain why some patient education materials are more readable than others and how that information might be used prospectively to improve their design. I therefore recommend revision, with particular emphasis on extending the PCA analysis in this direction.

Response

The author sincerely thanks the reviewer for this constructive direction and undertook the requested focused extension. The revised manuscript now uses the strong readability PCA result as the basis for individual webpage PC1 scoring, a new PC1 distribution figure, continuous feature-PC1 analyses with 95% confidence intervals, paired linguistic comparisons with effect sizes, within-category examination of diagnosis/treatment resources at the favorable and unfavorable extremes of PC1, and an exploratory feature-level PCA. Component scores from the retained feature-level PCA were also directly evaluated in relation to the original readability PC1 and page category. The feature-derived linguistic-complexity component was strongly associated with the original readability PC1 ($r = 0.946$, 95% CI: 0.895–0.972; $p < 0.001$) and was significantly higher for diagnosis/treatment materials than overview materials (0.507 vs.

–0.507; mean difference = 1.014, 95% CI: 0.532–1.496; $t(17) = 4.40$, $p < 0.001$; Cohen's $d_z = 1.05$). These extended analyses allowed the manuscript to move beyond documenting elevated readability levels toward identifying sentence- and lexical-complexity characteristics associated with greater textual difficulty. The Abstract, Materials and Methods, Results, Discussion, Limitations, recommendations, and Conclusion were revised accordingly while maintaining cautious interpretation appropriate to the sample size and study design.

Abstract, page 1; Materials and Methods, pages 4–5; Results, pages 9–14; Discussion/Limitations, pages 15–16; Conclusion, page 17.

Response to Reviewer

We thank the reviewer for the careful review of our revised manuscript and for the constructive suggestions. We appreciate the reviewer's assessment that the central scientific concerns have been addressed. We have made additional revisions to clarify the interpretation of the analyses, strengthen methodological transparency and reproducibility, and avoid overstatement. Our responses to each remaining comment are provided below.

Comment 1. The explanatory analysis is partly mathematically circular

Reviewer Comment:

The original readability formulas are constructed principally from sentence length, word length, character counts, and syllable counts. It is therefore expected that PC1 derived from those formulas will correlate strongly with mean sentence length, long-sentence frequency, mean word length, and multisyllabic-word frequency. Similarly, the very strong correlation between the feature-derived component and readability PC1 is not independent validation because both components are largely constructed from the same underlying properties.

Response:

Thank you for this important clarification. We agree that these associations are partly expected because conventional readability formulas incorporate sentence length and word-, character-, or

syllable-based measures. We revised the Discussion to explicitly state that the strong associations between the quantified textual features and readability PC1 are partly expected by mathematical construction. We also clarified that the strong correlation between the feature-derived linguistic-complexity component and readability PC1 should not be interpreted as independent validation of a separate linguistic mechanism. Instead, the extended analysis is now presented as identifying formula-relevant and potentially modifiable textual characteristics that varied across the sampled webpages.

Location of revision: Discussion, page 17, paragraph beginning “Because conventional readability formulas...”

Comment 2. “Similarly complex clinical information” is not demonstrated

Reviewer Comment:

The authors did not measure content coverage, clinical detail, accuracy, or completeness. Belonging to the same broad content category does not establish equivalent informational complexity.

Response:

We agree and revised the manuscript accordingly. We removed language implying that the clinical information itself was demonstrated to be similarly complex. The Results now describe the compared materials as “resources addressing diagnosis and/or treatment topics” and as “resources within the same broad diagnosis/treatment content category.” We also revised the Discussion to describe the observed differences in terms of measurable characteristics of how the information was written, including sentence and lexical complexity, rather than attributing those differences to unmeasured clinical-content complexity.

Location of revision: Results, pages 12 and 15; Discussion, page 16.

Comment 3. The institutional ANOVA is weak and overinterpreted

Reviewer Comment:

With only two webpages per institution and substantial content-category differences, the one-way institutional ANOVA has limited ability to detect institutional differences. The statement should avoid treating a nonsignificant result as evidence of equivalence.

Response:

We agree. We revised the Results to state that the analysis “did not detect statistically significant differences in readability among institutions,” rather than interpreting the nonsignificant findings as evidence of equivalent readability. We also added a limitation explaining that each institution contributed only two webpages representing different content categories and that the institution-level ANOVA therefore had limited statistical power and should not be interpreted as evidence of equivalent readability among institutions.

Location of revision: Results, page 10; Discussion/Limitations, page 17.

Comment 4. Reproducibility needs strengthening

Reviewer Comment:

The manuscript should report the precise URLs and access date, which sections were combined for overview datasets, the software or code used for textual-feature extraction, handling of headings, bullets, abbreviations, hyphenated terms, decimals, and fragments, whether PCA used a correlation or covariance matrix, how component scores were calculated, and preferably provide cleaned text, derived features, or code.

Response:

Thank you. We expanded the Materials and Methods section and Supplementary File 1 to improve methodological transparency and reproducibility. Supplementary File 1 now provides the source URLs for all 36 included patient education materials, their page classifications, and the recorded access period of February 2026. For institutions in which overview and diagnosis/treatment content appeared on the same webpage, the manuscript and supplement clarify that the relevant sections

were separated by content category and analyzed independently. For institutions with separate webpages, the overview and diagnosis/treatment pages were analyzed as distinct sources. The revised Methods also identify Microsoft Excel, Jamovi version 2.7.2, the Rj Editor, and R-based procedures used in the textual-feature analyses. We added operational details for sentence segmentation, word tokenization, syllable estimation, acronym identification, and MATTR-50 calculation. Complete sentences were defined using sentence-final punctuation (. ! ?), while headings and fragments without sentence-final punctuation were excluded. Alphabetic tokens containing internal apostrophes or hyphens were treated as words, and punctuation was excluded from mean word-length calculations. Syllable counts were based on the CMU Pronouncing Dictionary when available, with a deterministic vowel-group heuristic used for out-of-vocabulary words. Acronyms were defined as all-capital alphabetic abbreviations containing at least two letters, and MATTR-50 was calculated using consecutive 50-word windows. We also clarified that no separate manual rules were applied specifically to bullet markers or decimal values beyond the preprocessing and tokenization procedures implemented in the Rj Editor analysis. We further clarified that PCA was performed using the correlation matrix of the standardized readability variables. Component scores were generated from the retained PCA solution in Jamovi and were used without a custom author-defined scoring algorithm.

The exact subsection headings used for every overview dataset were not recorded during data collection and therefore cannot be reconstructed reliably without introducing retrospective assumptions. We have instead clarified the content-selection procedure used during data collection: when a single overview page contained limited material, multiple non-treatment sections, such as signs, risk factors, prevention, and general melanoma information, were combined to create the overview dataset, while treatment-specific material was analyzed separately.

For institutions in which the same webpage supplied both overview and diagnosis/treatment content, Supplementary File 1 now explicitly identifies those shared-source cases and states that the relevant sections were separated by content category and analyzed independently.

Location of revision: Materials and Methods, pages 3–6; Supplementary File 1.

Comment 5. One threshold statement should be made more precise

Reviewer Comment:

The manuscript states that every material exceeded the sixth-grade readability recommendation, although different formulas do not always produce the same classification. The metric used for this statement should be identified.

Response:

We agree and revised the Results to identify the specific readability metric used for this statement. The manuscript now states: “According to FKGL, all 36 materials were written above a sixth-grade reading level.” This avoids implying that all readability formulas produced the same classification.

Location of revision: Results, page 6.

Additional Comment Regarding the Paired Structure of the Regression

Reviewer Comment:

The paired structure of the data should be acknowledged or corrected in the regression analysis.

Response:

We agree and added an explicit limitation regarding the paired structure of the data. Because each institution contributed paired overview and diagnosis/treatment webpages, the observations were not fully independent at the institutional level. The revised Discussion now states that the multivariable regression did not explicitly model this within-institution correlation and that its estimates should therefore be interpreted as exploratory.

Location of revision: Discussion/Limitations, page 17.

We thank the reviewer again for these thoughtful and constructive comments. We believe that these revisions have improved the clarity, methodological transparency, and interpretation of the manuscript, and we appreciate the opportunity to further strengthen the work.

Thank you for addressing my comments. Please re-write the diagnosis/treatment materials per your discovered recommendation and compare scores. You can use an AI LLM such as chatgpt or equivalent to rewrite. You can have your mentor independently adjudicate content fidelity/preservation. Please acknowledge the LLM in the acknowledgement section. A formal review follows:

At present, the study establishes an already predictable result: diagnosis/treatment materials are harder to read than overview materials, largely because they contain longer sentences and more multisyllabic words. The manuscript recommends shorter sentences and simpler words but does not demonstrate that these changes can actually reduce readability scores while preserving the necessary medical information.

A much stronger study would convert the descriptive analysis into a controlled rewriting experiment.

Proposed intervention

The authors could rewrite all 18 diagnosis/treatment materials using a prespecified protocol:

- retain every clinically essential fact, warning, treatment option, numerical value, and recommendation;
- preserve the meaning of medical terms and define terms that cannot be simplified;
- shorten sentences, targeting approximately 15 words per sentence;
- reduce sentences longer than 20 words;
- replace unnecessarily complex words with shorter plain-language alternatives;
- use active voice and direct constructions;
- split dense paragraphs;
- expand or define acronyms at first use;
- use bullets for genuine lists;
- avoid deleting content merely to improve a formula score.

They would then calculate the same readability scores before and after rewriting.

Appropriate design

Each original diagnosis/treatment text would be paired with its rewritten version:

A negative difference for grade-level measures would indicate improvement. For Flesch Reading Ease, a positive difference would indicate improvement.

The primary endpoint should be prespecified—for example, change in FKGL. The remaining formulas should be sensitivity measures rather than five independent primary outcomes.

They should report:

- mean and median before-versus-after scores;
- mean paired change with 95% confidence interval;
- paired (t)-test or Wilcoxon signed-rank test;

Cohen's (d_z);

- percentage reaching the sixth-grade and eighth-grade targets;
- word-count change;
- preservation of clinically essential information;
- inter-rater agreement for content-preservation assessment.

A paired plot connecting each original and rewritten webpage would be far more informative than another PCA plot.

The essential safeguard: content fidelity

Merely lowering formula scores is easy. Authors can shorten words or delete medically important information and obtain a better score. That would not establish a better patient-education resource.

Each rewritten text therefore needs blinded evaluation by a medically knowledgeable reviewer (and preferably two reviewers if available) using a checklist derived from the original content. The reviewer should assess:

- factual accuracy;
- retention of essential information;
- preservation of risks and warnings;
- preservation of treatment alternatives;
- absence of newly introduced errors;
- absence of misleading simplification;
- actionability and understandability.

Using PEMAT for understandability and actionability would add information that readability formulas alone cannot provide. DISCERN could be considered where treatment-choice quality is relevant.

OPTIONAL : Ideally, the authors can also test the texts with patients or lay readers. Participants could be randomized to read either the original or rewritten version and then complete:

- objective comprehension questions;
- recall questions;
- perceived difficulty ratings;
- confidence or decisional-clarity measures;
- time required to read the material.

That would determine whether improved formula scores produce genuine improvements in understanding.

Please respect copyright

The authors can rewrite the text for research purposes and clearly label the versions as investigator-generated adaptations. They should not imply that the institutions approved or adopted the revisions. Full reproduction of website text may also raise copyright concerns; the manuscript may need to present excerpts, transformation rules, or supplementary adaptations within applicable permissions rather than republishing entire pages indiscriminately.

A stronger and more generalizable alternative would be to derive a rewriting protocol using part of the dataset and test it on held-out materials:

Use perhaps 12 diagnosis/treatment texts to develop the protocol.

Freeze the rewriting rules.

Apply them without further tuning to the remaining six texts.

Evaluate whether readability improves in this held-out set.

Without this separation, the protocol may simply be optimized to the same materials used to identify the problematic features.

The manuscript's new question would become:

Can a prespecified, evidence-based plain-language intervention reduce the measured textual complexity of melanoma diagnosis and treatment materials while preserving clinically essential content?

That question is much more consequential than asking whether several readability formulas agree.

A suitable revised title would be:

Improving the Readability of Online Melanoma Diagnosis and Treatment Materials: A Paired Plain-Language Rewriting Study

The central conclusion could then be causal within the tested texts: the defined rewriting procedure reduced calculated textual difficulty.

Author inquired by email of this was within scope and was required for acceptance. Editor checked with reviewer and replied that it was not. Accepted.

