Can Generative Artificial Intelligence Enhance Health Literacy About Lateral Epicondylitis?

Salvatore Capotosto, Michael Joseph Miskiewicz, Christian J Leonardo, Kenny Ling, Dorian Cohen[1], David E Komatsu[2], Edward D Wang[3]
[1]Stony Brook University Hospital, [2]Stony Brook University, [3]Stony Brook Univ Schl of Med

INTRODUCTION:

Patient educational materials (PEMs) about lateral epicondylitis (LE) from top orthopaedic institutions are often written above the 8th grade level recommended by the Centers for Disease Control and Prevention (CDC) and the National Institutes of Health (NIH). Our study aims to evaluate the ability of generative artificial intelligence (AI) to enhance readability of PEMs about LE from the top 25 US orthopaedic institutions.

METHODS:

PEMs about LE from the top twenty-five ranked orthopaedic institutions from the 2022 U.S. News & World Report Best Hospitals Specialty Ranking were gathered. For each set of PEMs, readability was assessed using seven distinct formulas: Gunning Fog, Flesch-Kincaid Grade Level, Coleman-Liau Index, Simple Measure of Gobbledygook (SMOG) Index, Automated Readability Index, Linsear Write Formula, and FORECAST Readability Formula. ChatGPT Plus (version 4.0) was then instructed to rewrite PEMs on LE from these institutions to comply with CDC and NIH recommended guidelines. Specifically, ChatGPT was prompted to: (1) limit the total number of polysyllabic words to less than 30, (2) limit sentences to less than 10 words, (3) limit paragraphs to less than 5 sentences, (4) eliminate as much medical jargon without compromising accuracy, (5) when eliminating medical jargon is not possible, provide a brief explanation of the relevant concept, and (6) overall, rewrite this as if you were speaking to an eighth grader. Readability scores were calculated for the original and rewritten PEMs, and paired t-tests were used to determine statistical significance.

RESULTS:

Twenty-two of the initial twenty-five orthopaedic institutions contained educational material related to LE on their websites. In the unedited PEM cohort, only six institutions obtained average readability scores below the 8th grade reading level. The average reading grade level of all institution's unedited PEMs was 9.81 ± 1.76, and the average word count was 600.68 ± 409.44 words. Following ChatGPT's edits to the original PEMs, all twenty-five rewritten PEMs obtained average readability scores below the 8th grade reading level. The total average readability score for all rewritten PEMs was 6.12 ± 0.97, with an average word count of 253.96 ± 100.76 (Figure 1-2). By utilizing ChatGPT to rewrite the original PEMs, a reduction of 3.70 ± 1.84 ($p<0.001$) reading grade levels and 346.72 ± 364.63 ($p<0.001$) words was achieved (Table 1). The senior orthopaedic resident and attending orthopaedic surgeon validated the accuracy of the information in the PEMs rewritten by ChatGPT.

DISCUSSION AND CONCLUSION:

In conclusion, our study showcases the capacity of generative AI to enhance PEMs for LE, ensuring compliance with CDC and NIH guidelines. Hospital administrators and surgeons should explore AI's potential in improving PEMs' readability and promoting broader access to health knowledge.

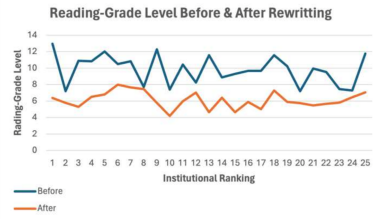

Figure 1. The average readability scores of patient education materials related to lateral epicondylitis from twenty-five of the top nationally-ranked orthopedic institutions, before and after generative AI-assisted editing.

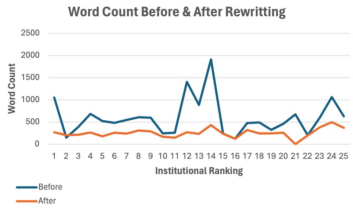

Figure 2. The average word count of patient education materials related to lateral epicondylitis from twenty-five of the top nationally-ranked orthopedic institutions, before and after generative AI-assisted editing.

	Mean Reduction	Standard Deviation	95% Confidence Interval		p-value
			Lower	Upper	
Readability Score	3.70	1.84	2.93	4.46	<0.001
Word Count	346.72	364.63	196.21	497.24	<0.001

Table 2: Results for paired t-test to determine significance between (a.) Average readability scores, and (b.) Total word count of the original and rewritten PEMs.