

Improving the Readability of the Pediatric Orthopaedic Society of North America Trauma Education Materials Using ChatGPT-4o

Diane I Ghanem, Elyette M Lugo, Raisa Rauf, James R Ficke, Rushyuan Jay Lee, Babar Shafiq, Ismat Ghanem
INTRODUCTION:

Previous studies have shown that online pediatric orthopaedic trauma education materials often exceed the recommended sixth-grade reading level, potentially limiting patient comprehension. This study aimed to determine whether large language model (LLM) technology, specifically ChatGPT-4o, could effectively improve the readability of the Pediatric Orthopaedic Society of North America (POSNA) trauma materials.

METHODS:

All nine publicly available trauma education articles from the POSNA website were included. Each article's baseline readability was assessed using the Flesch-Kincaid Grade Level (FKGL) and Flesch Reading Ease (FRE) metrics by two independent examiners. Articles were then provided to ChatGPT-4o with structured prompts to target a sixth-grade reading level: "You are an expert in medical writing and patient education. Please re-write the following text at a 6th-grade reading level (Flesch-Kincaid Grade Level between 6 and 7) by simplifying sentences and using more accessible vocabulary while preserving the original meaning." Post-revision, the FKGL and FRE scores were recalculated. Paired two-tailed t-tests were performed to evaluate differences between the original and ChatGPT-revised versions. Interobserver reliability was assessed using intra-class correlation coefficients (ICCs).

RESULTS:

ChatGPT-4o significantly improved readability across all POSNA trauma education materials. The mean FKGL significantly decreased from 9.44 ± 1.88 to 5.29 ± 0.77 ($p < 0.001$), achieving the recommended sixth-grade level. Similarly, the FRE scores significantly improved from 61.44 ± 10.02 to 80.84 ± 5.10 ($p < 0.001$), shifting materials from "fairly difficult" to "easy" to read. The clavicle and pelvic avulsion fracture articles were initially the most difficult to read (FKGL 11.9 and 12.4, and FRE 54.4 and 40.6, respectively), while the ankle and growth plate articles were the easiest (FKGL 7.05 and 7.35, and FRE 71.45 and 71.25, respectively). After revision, all articles met readability benchmarks, with the clavicle fracture article showing the greatest absolute improvement in FKGL (-7.7). Reviewer agreement was excellent ($ICC > 0.90$).

DISCUSSION AND CONCLUSION:

ChatGPT-assisted revision substantially improved the readability of pediatric orthopaedic trauma education materials from POSNA, achieving the recommended sixth-grade reading level as per the established health literacy standards. LLMs, when guided by expert review, may offer a powerful, scalable solution to bridge the gap between clinical accuracy and public accessibility.

