

# Optimizing the Readability of Orthopaedic Trauma Patient Education Materials Using ChatGPT-4

Oscar Covarrubias<sup>1</sup>, Diane Ghanem<sup>2</sup>, Christopher Murdock<sup>2</sup>, Babar Shafiq<sup>2</sup>  
<sup>1</sup>Johns Hopkins University, School of Medicine, Baltimore, MD, <sup>2</sup>The Johns Hopkins Hospital, Baltimore, MD  
dghanem1@jh.edu

**Disclosures:** No relevant disclosures.

**INTRODUCTION:** ChatGPT is an advanced language AI tool designed to understand and generate human-like text. The aim of this study is to assess the ability of ChatGPT-4 to re-write orthopaedic trauma patient education materials at the recommended 6<sup>th</sup>-grade level. Amidst this transformative shift, the aim of this study is to assess ChatGPT's performance on the orthopaedic surgery in-training examination (OITE).

**METHODS:** Two independent reviewers accessed ChatGPT-4 (chat.openai.com) and gave identical instructions to simplify the readability of provided text to a 6<sup>th</sup>-grade level. All trauma-related articles by the Orthopaedic Trauma Association (OTA) and American Academy of Orthopaedic Surgeons (AAOS) were sequentially provided. The academic grade level was determined using the Flesh-Kincaid Grade Level (FKGL) and Flesch Reading Ease (FRE). Paired t-tests and Wilcoxon-rank sum tests were used to compare the FKGL and FRE between the ChatGPT-4 revised and original text. Inter-rater correlation coefficient (ICC) was used to assess variability in ChatGPT-4 generated text between the two reviewers.

**RESULTS:** ChatGPT-4 significantly reduced FKGL and increased FRE scores in the OTA (FKGL: 5.7±0.5 compared to the original 8.2±1.1, FRE: 76.4±5.7 compared to the original 65.5±6.6,  $p < 0.001$ ) and AAOS articles (FKGL: 5.8±0.8 compared to the original 8.9±0.8, FRE: 76±5.5 compared to the original 56.7±5.9,  $p < 0.001$ ). On average 14.6% of OTA and 28.6% of AAOS articles required at least two revisions by ChatGPT-4 to achieve a 6<sup>th</sup>-grade reading level. ICC demonstrated poor reliability for FKGL (OTA 0.24, AAOS 0.45) and moderate reliability for FRE (OTA 0.61, AAOS 0.73).

**DISCUSSION:** This study provides a novel, simple and efficient method using language AI to optimize the readability of patient education content which may only require the surgeon's final proofreading. This method would likely be as effective for other medical specialties.

**SIGNIFICANCE/CLINICAL RELEVANCE:** By providing a novel, simple and efficient solution using the language AI model ChatGPT Plus, this study bridges the ubiquitous literacy-readability gap across orthopaedic patient education materials, which are often too complex for patients to understand. This tool can be used to assist all professionals in the field of orthopaedic surgery, including the AAOS and the OTA, to tailor patient education materials at the recommended 6<sup>th</sup>-grade level, thus enhancing patient comprehension and improving patient outcomes.