Assessing and Improving Readability of AAOS Spanish-Language Patient Education Materials Using Artificial Intelligence

Rodnell Busigó Torres, Mariana Restrepo, Brocha Stern¹, Bernardo Yahuaca, Rafael Buerba Siller², Ivan Alonzo Garcia³, Victor H Hernandez, Ronald Anthony Navarro

¹Icahn School of Medicine At Mount Sinai, ²Banner Health Center Plus, ³Kaiser Permanente INTRODUCTION:

More than 41 million people in the USA speak Spanish, representing one of the fastest-growing populations in the country. Hispanics accounted for over half of the total U.S. population growth from 2010 to 2022. Non-English speaking patients often experience poorer health outcomes, potentially due to language barriers that hinder patient education. Patient educational materials (PEMs) in orthopedics are essential for informing patients about diagnoses, surgical procedures, recovery processes, and post-operative care. However, the availability of these materials in Spanish is limited. The American Academy of Orthopaedic Surgeons (AAOS) recently translated PEMs into Spanish to assist orthopaedic surgeons in educating their Spanish-speaking patients. However, the readability levels of these Spanish PEMs are unknown. Previous studies have shown that English-language PEMs often exceed the recommended reading level (below 6th grade), leading to misunderstandings and barriers to effective patient care. This issue is particularly significant for the Hispanic population, which has the lowest health literacy rates. This study aimed to examine the readability levels of AAOS Spanish-language PEMs and evaluate if Artificial Intelligence (AI) can improve these readability levels.

METHODS:

We extracted all Spanish-language PEMs from the AAOS website, excluding articles related to COVID-19. We calculated the Fernandez-Huerta readability score, Spanish SMOG readability formula, and word count for each PEM. PEMs with a grade reading level above the 6th grade were input into ChatGPT-4 with the instruction "traduce al nivel de lectura de quinto grado" (translate to a fifth-grade reading level). After conversion by ChatGPT, we reassessed the readability metrics and compared them to those of the original AAOS version using an independent t-test. Four Spanish speaking orthopaedic surgeons evaluated the accuracy and usefulness for patient education of the ChatGPT-converted PEMs compared to the original AAOS versions using 5-point Likert scales; a response of "agree" or "strongly agree" was used to reflect comparable quality of the conversion.

RESULTS:

We included a total of 77 PEMs. The median Fernandez-Huerta score was 58/100 (9th to 10th grade; 95% CI 56-59), and the median SMOG reading level was 10.4 (10th to 11th grade; 95% CI 10.1-10.8). None of the AAOS PEMs were below the recommended 6th-grade reading level. After undergoing ChatGPT conversion, readability levels improved: the Fernandez-Huerta score increased to 67/100 (7th to 8th grade; 95% CI 62-66; p<0.001), a 13.4% improvement, and the SMOG reading level decreased to 8.7 (8th to 9th grade; 95% CI 7.9-8.5; p<0.001), a 16.4% improvement. Regarding the Fernandez-Huerta score, only 18/77 AI conversions (23.4%) achieved the threshold of below a 6th-grade reading level. For the SMOG readability formula, only one AI conversion (1.3%) met the threshold. Additionally, AI conversions were also significantly shorter in word count compared to the AAOS versions (239 vs 1,362 words; p<0.001), an 82.4% reduction. According to the orthopaedic surgeons' assessments, only 52.0% of AI conversions had comparable accuracy and only 55.8% had comparable usefulness for patient education compared with the original AAOS versions. DISCUSSION AND CONCLUSION:

This study highlights the inadequacies in the readability levels of AAOS Spanish PEMs, none of which met the recommended below 6th-grade reading level. This issue can further exacerbate health disparities among Spanish-speaking patients. Additionally, while previous literature has suggested that ChatGPT can improve readability levels in English, this study demonstrates limited success for readability of Spanish-language documents. While readability was improved, it remained difficult, and concerns were noted regarding accuracy and usefulness of the converted materials. The substantial reduction in word count may indicate that ChatGPT omitted important details necessary for patient education. This underscores that ChatGPT-4, in its current form, is not yet a reliable tool for enhancing the readability of Spanish PEMs without compromising accuracy. Future research should focus on developing methods to improve readability without sacrificing the quality of information.