



# Generative AI to Improve Readability of Surgical Patient Education Materials to Match U.S. Average Reading Level



Nguyen Tran, BS, Ainsley Durning, BS, Abigail Hegwood, MS, John Ashurst, DO, EdD, MS & the MIRAGE Research Group  
Arizona College of Osteopathic Medicine, Midwestern University

## Introduction

Patient educational materials (PEMs) often present above the National Institute of Health (NIH) recommended sixth grade reading level. With the average reading level of the U.S. population at an eight-grade level, this mismatch can lead to decreased health literacy and compliance with post-operative instructions. Low health literacy has been associated with poorer surgical outcomes. Utilization of AI platforms increases availability of health education and can be a great tool to improve health literacy.

## Objective

To evaluate the performance of Artificial Intelligence (AI) programs in rewriting PEMs to match the health literacy of an average U.S. surgical patient.

## Methods

Eight PEMs (appendectomy, laparoscopic cholecystectomy, colectomy, adult ventral hernia repair, adult and pediatric inguinal/femoral hernia repair, adult and pediatric umbilical hernia repair) were downloaded from the American College of Surgeons and separated into subsections. ChatGPT, Google AI, and Claude AI were instructed to rewrite text to a sixth-grade reading level while preserving the accuracy and completeness of the medical information. The Flesh-Kincaid calculator, a commonly used readability assessment tool, was utilized as the primary indicator of PEM readability. The Flesh-Kincaid equations use a combination of total number of words, characters, and sentences. The Flesh-Kincaid Grade Level (FKGL) calculator was used to assess text readability and the Flesch Reading Ease Score (FRES) calculator was used to assess understanding. Data analysis was performed using a Kruskal Wallis test and post-hoc Dunn test. Statistical significance was set at  $p \leq 0.05$ .

FKES	$206.835 - 1.015 \times (\text{Total Words} / \text{Total Sentences}) - 84.6 \times (\text{Total Syllables} / \text{Total Words})$
FKGL	$0.39 \times (\text{Total Words} / \text{Total Sentences}) + 11.8 \times (\text{Total Syllables} / \text{Total Words}) - 15.59$

FKES	Estimated FKGL
90 - 100	5th grade
80 - 90	6th grade
70 - 80	7th grade
60 - 70	8th and 9th grade
50 - 60	10th to 12th grade
30 - 50	College
0 to 30	College graduate

## Results

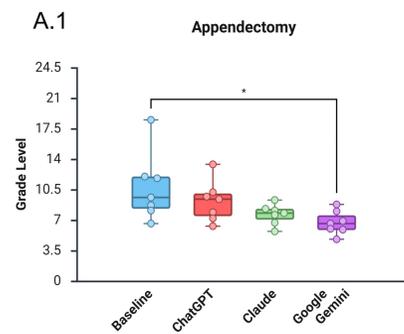


Figure A.1 is Appendectomy Grade Level: Grade level required to read a piece of text easily. There is statistical significance between Baseline vs Google Gemini (\* is  $p < 0.05$ ).  
Figure A.2 is Appendectomy Ease Score: Higher score indicates an easy-to-understand text. There is statistical significance between Baseline vs Google Gemini (\*\* is  $p < 0.01$ ).

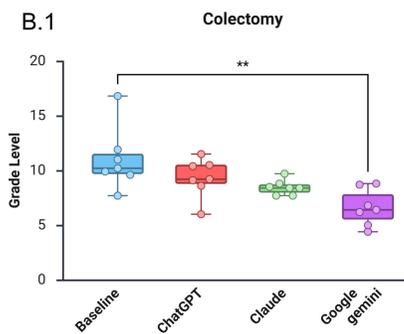
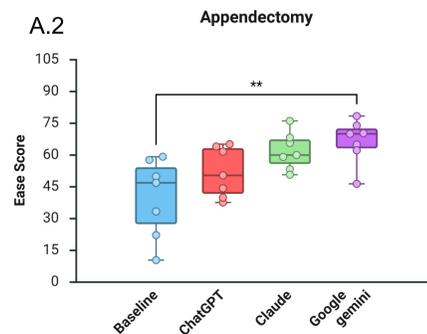


Figure B.1 is Colectomy Grade Level: Grade level required to read a piece of text easily. There is statistical significance between Baseline and Google Gemini (\*\* is  $p < 0.01$ ).  
Figure B.2 is Colectomy Ease Score: Higher score indicates an easy-to-understand text. There is statistical significance between Baseline and Google Gemini (\*\* is  $p < 0.001$ ).

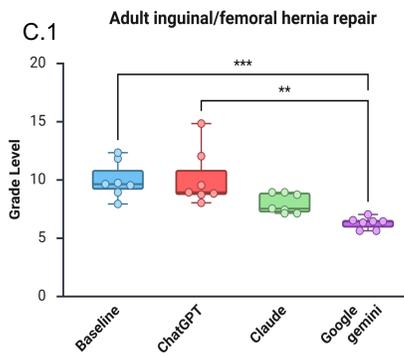
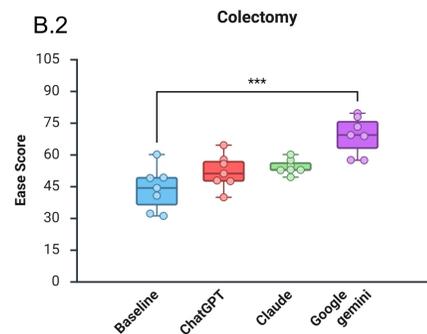


Figure C.1 is Adult Inguinal/Femoral hernia repair Grade Level: Grade level required to read a piece of text easily. There is statistical significance between Baseline vs Google Gemini, and ChatGPT vs Google Gemini (\*\* is  $p < 0.01$ ; \*\*\* is  $p < 0.001$ ).  
Figure C.2 is Adult Inguinal/Femoral hernia repair Ease Score: Higher score indicates an easy-to-understand text. There is statistical significance between Baseline and Google Gemini, and Chat GPT vs Google Gemini (\*\* is  $p < 0.01$ ; \*\*\* is  $p < 0.001$ ).

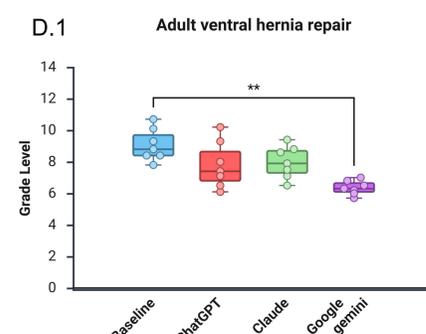
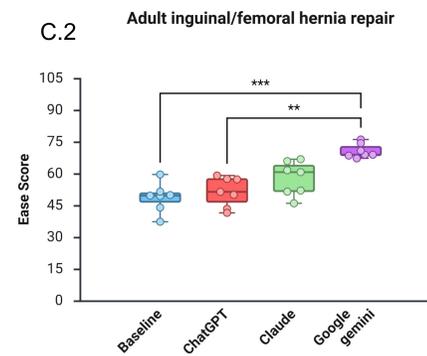


Figure D.1 is Adult ventral hernia repair Grade Level: Grade level required to read a piece of text easily. There is statistical significance between Baseline vs Google Gemini (\*\* is  $p < 0.01$ ).  
Figure D.2 is Adult ventral hernia repair Ease Score: Higher score indicates an easy-to-understand text. There is statistical significance between Baseline vs Google Gemini (\* is  $p < 0.05$ ).

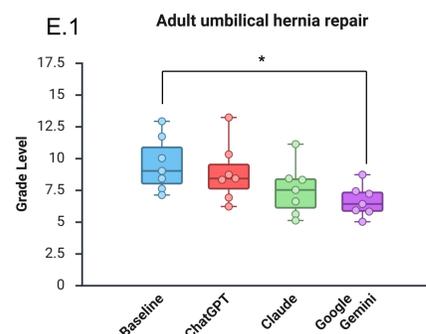
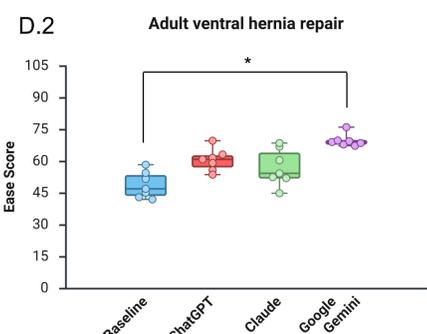


Figure E.1 is Adult umbilical hernia repair Grade Level: Grade level required to read a piece of text easily. There is statistical significance between Baseline vs Google Gemini (\* is  $p < 0.05$ ).  
Figure E.2 is Adult umbilical hernia repair Ease Score: Higher score indicates an easy-to-understand text. There is statistical significance between Baseline vs Google Gemini (\*\* is  $p < 0.01$ ).

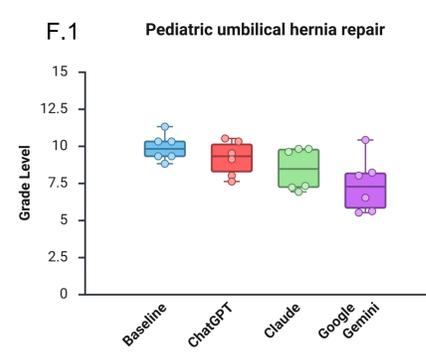
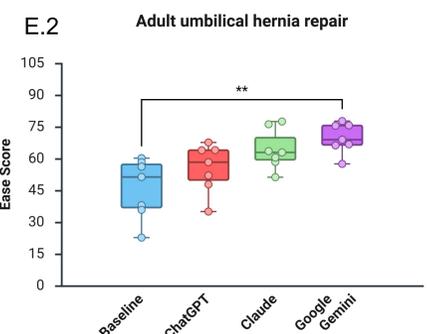
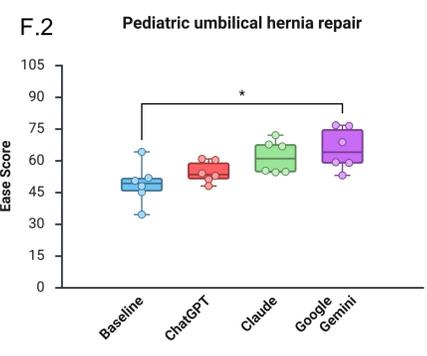


Figure F.1 is Pediatric umbilical hernia repair Grade Level: Grade level required to read a piece of text easily. There is statistical significance between Baseline and Google Gemini (\* is  $p < 0.05$ ).  
Figure F.2 is Pediatric umbilical hernia repair Ease Score: Higher score indicates an easy-to-understand text. There is statistical significance between Baseline and Google Gemini (\* is  $p < 0.05$ ).



## Discussion

PEMs provided by majority of high impact journals are written at a level significantly higher than the NIH recommended sixth grade level, negatively affecting health literacy of the general population. Medical decision making relies on comprehensive understanding. Google Gemini, as an adjunct with explanations from visits with providers, is a great alternative for the public to learn about their medical conditions and available elective surgical options.

Utilization of AI platforms in conjunction with instructions from providers can improve US health literacy as AI platforms continue to grow with new data. This project shows that Google Gemini can be a critical tool in improving patient health literacy for PEMs. Further research needs to be conducted to establish a concrete scientific method to improve the readability and understanding of PEMs.

## Conclusion

## References

- Burns, S. T., Amobi, N., Chen, J. V., O'Brien, M., & Haber, L. A. (2021). Readability of Patient Discharge Instructions. *Journal of General Internal Medicine*. <https://doi.org/10.1007/s11606-021-06988-y>
- Choudhry, A. J., Younis, M., Ray-Zack, M. D., Glasgow, A. E., Haddad, N. N., Habermann, E. B., Jenkins, D. H., Heller, S. F., Schiller, H. J., & Zielinski, M. D. (2019). Enhanced readability of discharge summaries decreases provider telephone calls and patient readmissions in the posthospital setting. *Surgery*, 165(4), 789-794. <https://doi.org/10.1016/j.surg.2018.10.014>
- Eltorai, A., Ghani, S., Adams, C., Born, C., & Daniels, A. (2014). Readability of Patient Education Materials on the American Association for Surgery of Trauma Website. *Archives of Trauma Research*, 3(1). <https://doi.org/10.5812/at.18161>
- Sahhar, M., Singh, M., Mehla, T., Laud, A., Nassar, J. E., Farias, M. J., Diebo, B. G., & Daniels, A. H. (2025). Lost in Translation: Preoperative Orthopaedic Education Materials Significantly Exceed Recommended Reading Levels. *JBJS Open Access*, 10(3). <https://doi.org/10.2106/jbjs.oa.25.00143>