

Seeing it in the Flesch: Comparing readability between AI-generated and human-written plain-language abstracts

32

David McMinn,^a Tomas Valena,^b & Walter Bender^c

^aSorcery, East Kilbride, UK
^bSorcery, Basel, Switzerland
^cSorcery, Washington, DC, USA



Background & aims

- There are both ethical and regulatory imperatives for developing plain-language summaries (PLSs) of scientific research that are understandable to a non-expert audience.¹
- Analyses of human-written PLSs show that they are “not so plain”, they are “difficult to read”, and they “require more effort” to write.^{2,3}
- Our approach to improving the quality, readability, and accessibility of these documents is through the use of artificial intelligence (AI).
- In this study, we compare readability scores and population-level accessibility of AI-generated abstract PLSs (aPLS) vs corresponding published (human-written) aPLSs.

Methods

1

Literature Search

- PubMed was searched using the term ‘Plain Language Summary’.
- Articles (n=50) were selected that contained aPLSs written by medical-writing professionals.
- The corresponding scientific abstracts (n=48) were identified for each aPLS (n=2 could not be located).

2

AI aPLS Generation

- An aPLS was auto-generated using AI for each scientific abstract.
- Two bespoke AI approaches were used, each based on several large language models and task-specific coding:
 - Approach 1: biobert, text-bison@001, bart-large-mnli, scibert
 - Approach 2: biobert, gpt-3.5-turbo, bart-large-mnli, scibert
- Process optimization steps were taken to ensure accurate source-content classification and discrepancy checks on AI-generated content.

3

Outcomes & Analysis

- Each aPLS was assessed on the following measures:
 - Readability metrics: Flesch-Kincaid, Automated Readability Index (ARI), Linsear-Write, and SPACHE
 - Accessibility: average readability score for each aPLS type plotted against population literacy distribution (PIAAC)
 - Reading time⁴
 - Subject matter expert (SME)-assessed quality: good, moderate, poor (assessed before and after process optimization)
- Readability metrics were compared across abstract type using ANOVA with pairwise comparisons and p values.

Results

- Future Oncology (n=23) and Neurodegen. Dis. Manag. (n=7) were the most frequently represented journals.
- aPLSs generated using the bespoke AI process with task-specific coding had significantly lower readability scores (i.e. were easier to read) than the human-written equivalents (Figure 1).
- Consequently, AI-generated aPLSs are accessible to a larger percentage of the population (~50%) than both scientific abstracts (~1%) and human-written aPLSs (~12%) (Figure 2).
- Both AI-generated aPLSs and human-written aPLSs had significantly lower reading times than the corresponding scientific abstracts (Figure 3).
- Using AI Approach 2, more than 75% of aPLSs were categorised as “Good” quality by SMEs (Figure 4).

Figure 1. Readability scores across aPLS types

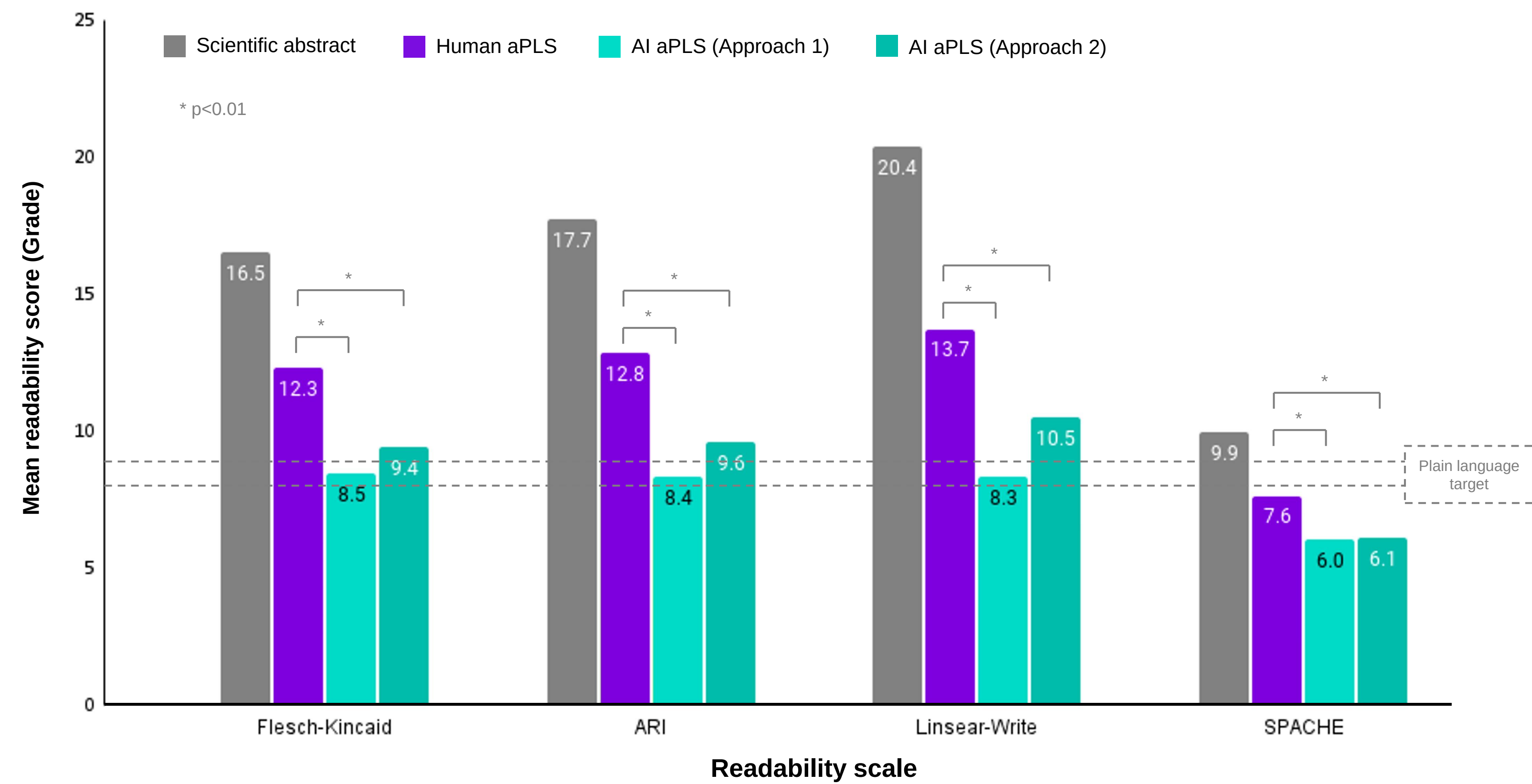


Figure 2. Population-level accessibility

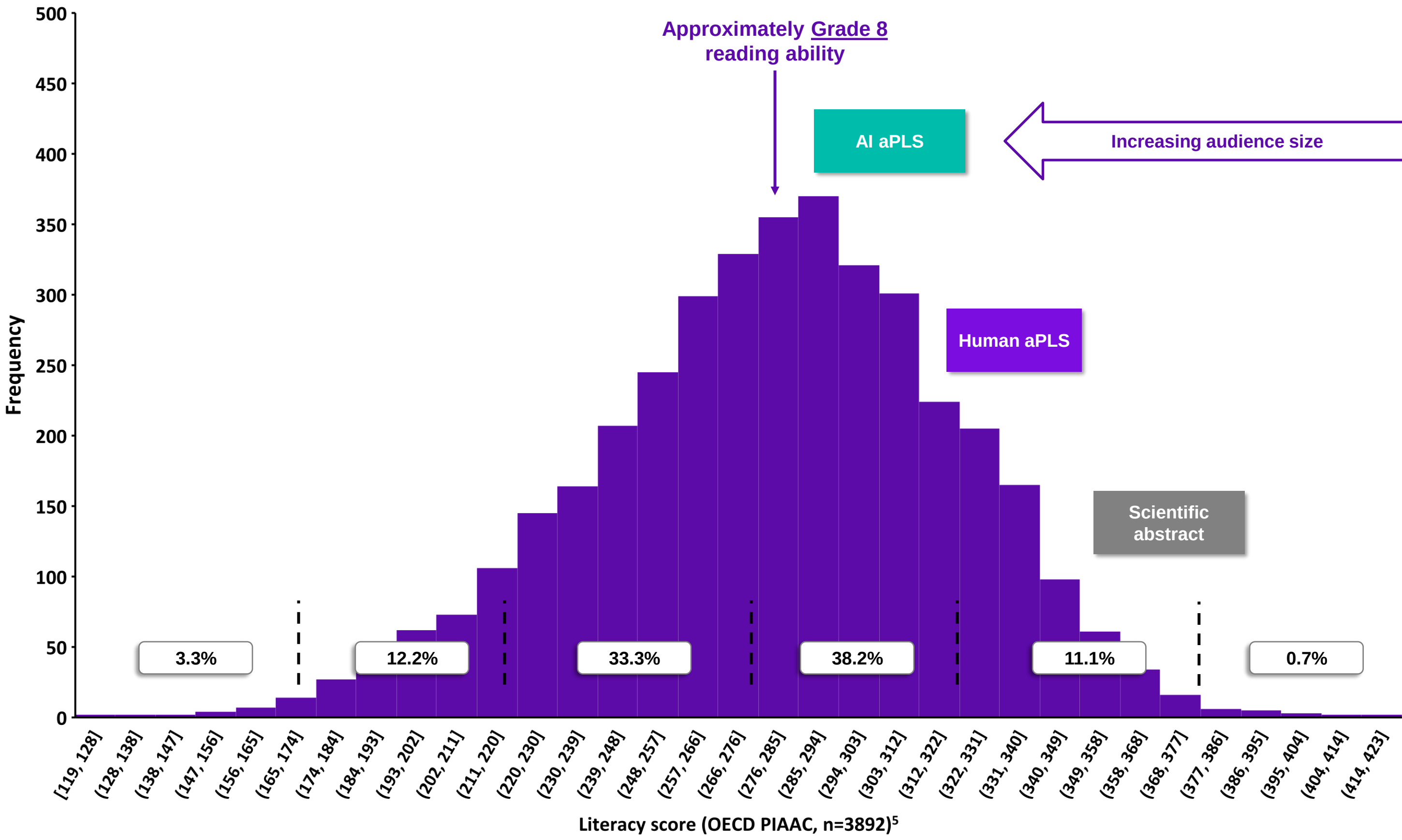


Figure 3. aPLS reading times

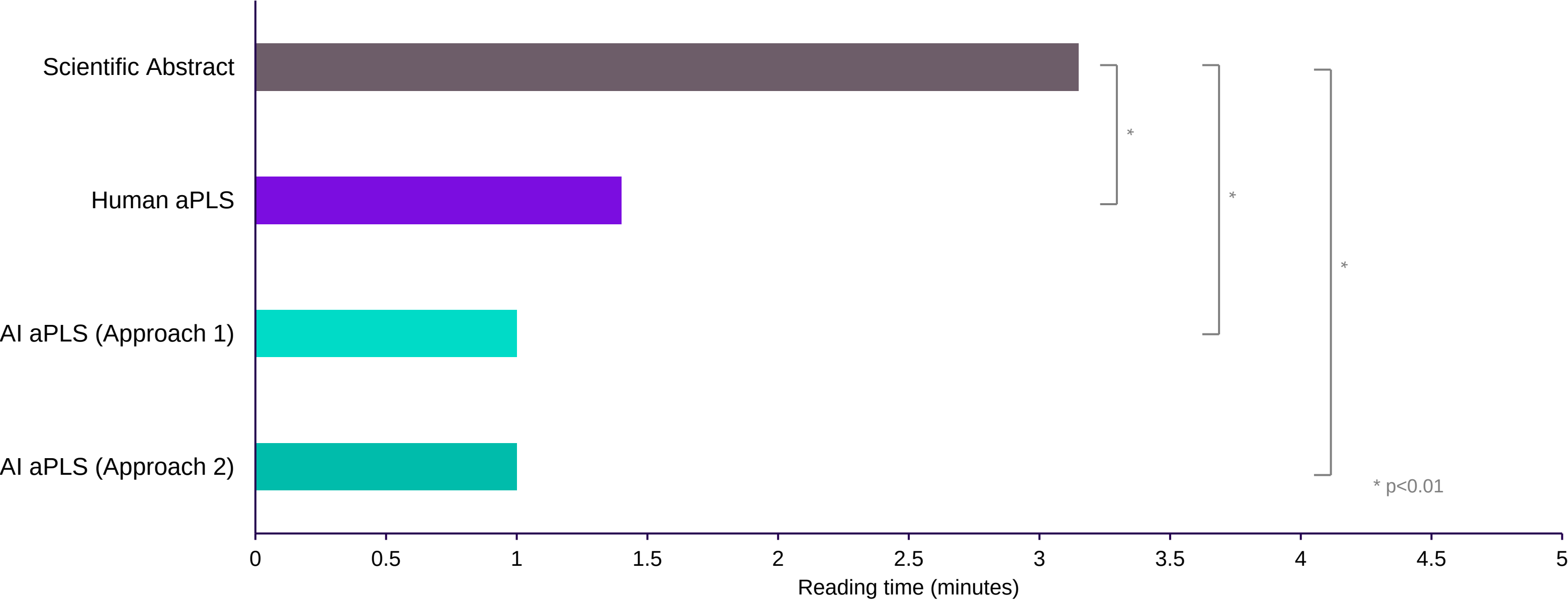
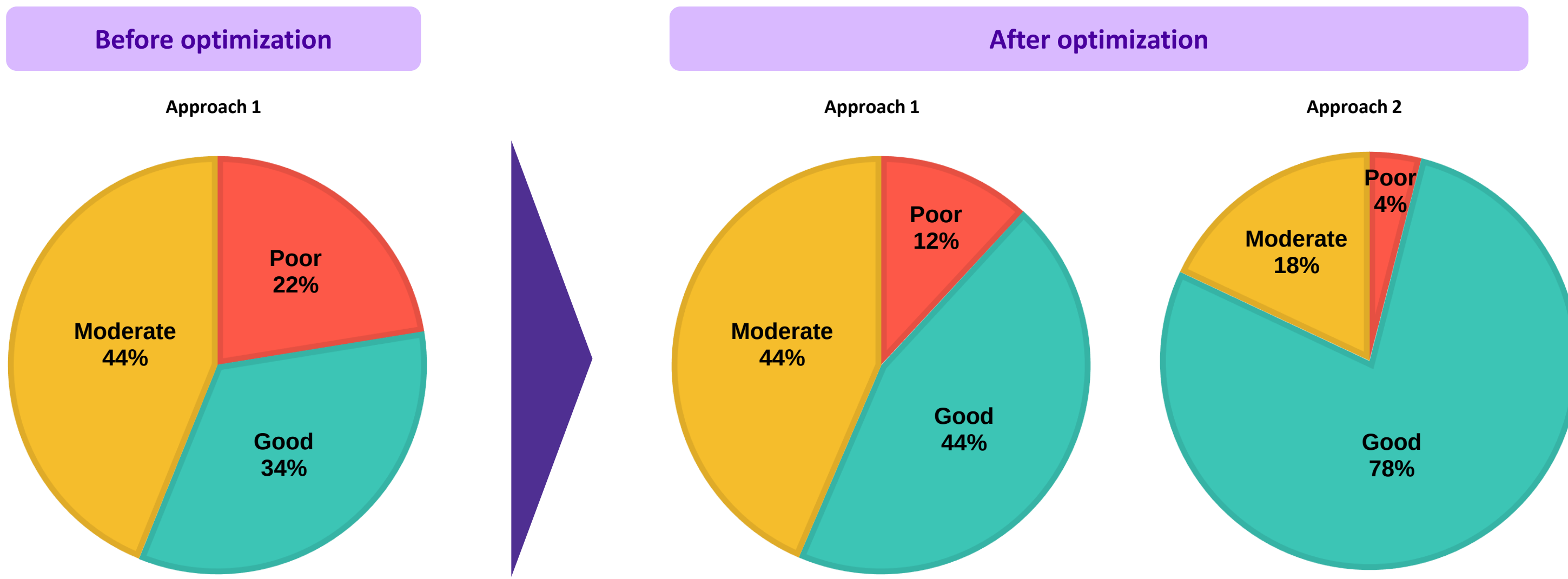


Figure 4. SME-assessed quality of AI-generated aPLSs



Conclusions

- Plain-language summaries of scientific content generated using bespoke AI processes and task-specific coding are easier to read and are more accessible than those written by medical-writing professionals.
- Research is ongoing to investigate the labour-saving potential of this approach for aPLS generation; a follow-up study includes end-user assessments.

References

¹European Commission Clinical Trials Expert Group. Good Lay Summary Practice. October 2021. ²Banic A, et al. BMC Medical Research Methodology 2022;22:240. ³Shiely F & Daly A. Journal of Clinical Epidemiology 2023;156:105-112. ⁴Demberg V & Keller F. Cognition 2008;109:193-210. ⁵US OECD PIAAC literacy data: <https://webfs.oecd.org/piaac/puf-data/csv/>

Disclosures

- David McMinn, Tomas Valena, and Walter Bender are employees of Sorcery.
- This poster is an encore of an oral presentation given at ISPOR EU 2023, Copenhagen.