

Investigating sentence reading in young adults with and without university education: An eye-tracking study

Kathy Conklin¹ and Rüdiger Thul²

¹ Centre for Research in Applied Linguistics, School of English, University of Nottingham

² School of Mathematical Sciences, University of Nottingham, Nottingham

While reading plays a vital role in modern society (Weiss et al., 2016), much of the research on reading has been carried out on children, populations with reading difficulties, and university students participating for course credit or payment with the goal of understanding basic reading processes, assessing reading development, and ascertaining the cognitive benefits of reading (Venezky & Sabatini, 2002; Krueger et al., 2023). Thus, while significant work has been done to understand reading in children and specialized adult populations, we have much less of an understanding of the potential variability in healthy adult readers across a broad cross-section of society for whom decoding and reading is automatized and where the primary aim of reading is to understand written texts and gain new information (Scagnelli et al., 2021).

The current study addresses this gap by investigating reading in unimpaired, matched native English-speaking populations with and without university education using authentic texts. We monitored the eye movements of participants without a university education ($n = 10$; mean age = 21) and a matched group of undergraduate university students ($n = 13$; mean age = 18.6) while they read 10 BBC articles (total words = 7,444). After reading an article, participants rated it for: informativeness, ease of reading, being well-written and being enjoyable.

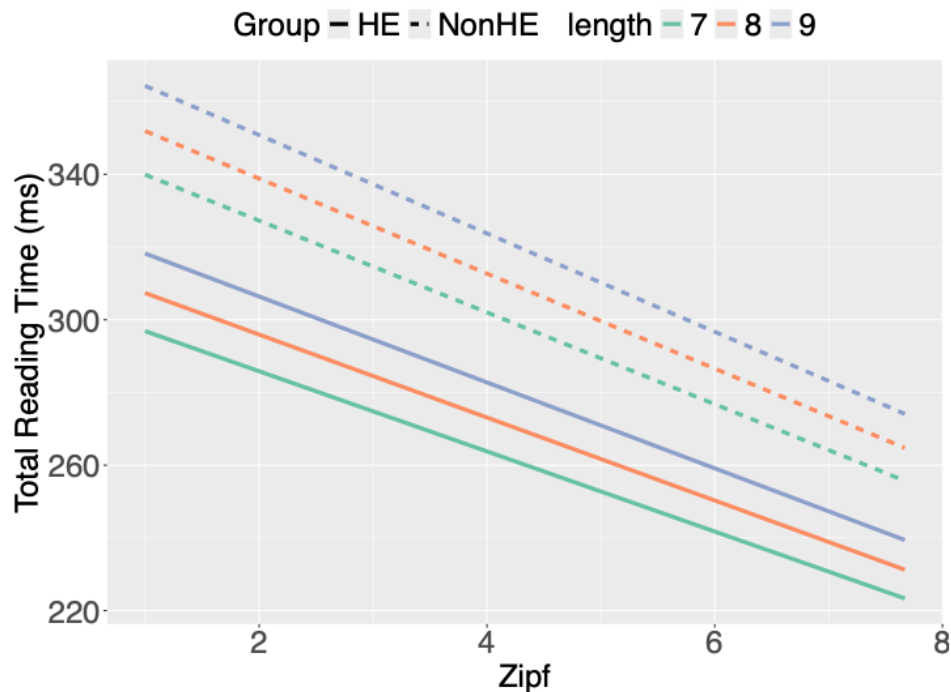
In terms of the rating data, there was a strong correlation between an article being perceived as easy to read and as being well written amongst participants without a university education ($\rho = 0.67$). However, this relationship was less evident in university student participants ($\rho = 0.28$).

When reading sentences from authentic texts, some words are read repeatedly in varied contexts, yielding different reading times for the same word within and across participants (e.g., the word “drinks” occurred 19-times in a single article). To tackle this heterogeneity, we employed generalized linear mixed models. First fixation duration and total reading time were both best captured by a distribution from the Box-Cox family, which are well adapted for skewed data with longer tails. Compared to participants with university education, participants without university education had longer and more variable first fixation durations ($\beta_1 = 0.094, t = 10.1$; $\beta_2 = 0.061, t = 10.82$) and longer, but less variable total reading time ($\beta_1 = 0.118, t = 9.88$; $\beta_2 = -0.046, t = -2.45$). Both group’s first fixation durations and total reading time decreased with increasing Zipf values ($\beta_3 = -0.02, t = -16.06$; $\beta_3 = -0.056, t = -32.8$), and so did variability in reading time. Put differently, more frequent words were read more quickly, and reading times became less dependent on the context. For articles that participants rated as challenging and less enjoyable, there was a significant interaction between word frequency and group, i.e., a greater processing costs for low frequency words in these articles for participants without a university education (see Figure 1). Figure 1 also illustrates a robust word length effect.

Our work demonstrates that it is critical to move beyond datasets from Western, Educated, Industrialized, Rich, and Democratic (WEIRD) populations to more fully understand the complexities that underpin natural reading in a more representative manner. Crucially, we show that less-WEIRD data is more heterogenous; participants without a university education exhibit more variability in their reading behavior as quantified through larger coefficients of variations. Notably, the emergent positively-skewed and long-tailed distributions from the Box Cox family point to a non-normality related to the context in which words are found. This indicates that

contexts are not symmetric, in that some contexts can increase reading times substantially more compared to the reduction of reading times in other contexts. Finally, gaining a greater understanding of how reading changes in context and between groups of readers may help writers/authors to make their texts more accessible for a wider audience.

Figure 1 Total reading time for selected word lengths as a function of word frequency (Zipf) in young readers with a university education (HE) and without (NonHE)



References

- Krueger, K. R., Desai, P., Beck, T., Wilson, R. S., Evans, D., & Rajan, K. B. (2023). Cognitive activity is associated with cognitive function over time in a diverse group of older adults, independent of baseline biomarkers. *Neuroepidemiology*, 57(4), 229-237. <https://doi.org/10.1159/000531208>
- Scagnelli, M., Della Beffa, F., & Santulli, F. (2021). Reading and Comprehension in Adulthood: a Training Programme. *Journal of Cognitive Enhancement*, 5, 25-34. <https://doi.org/10.1007/s41465-020-00182-4>
- Venezky, R. L., & Sabatini, J. P. (2002). Introduction to this special issue: Reading development in adults. *Scientific Studies of Reading*, 6(3), 217-220. https://doi.org/10.1207/S1532799XSSR0603_1
- Weiss, B., Knakker, B., & Vidnuanszky, Z. (2016). Visual processing during natural reading. *Scientific Reports*, 6, 1–16. <https://doi.org/10.1038/srep26902>