The role of experience with visual and linguistic factors on skipping behavior during reading naturalistic English texts in English L1 teen and adult readers

Anastasia Stoops and Jessica L. Montag (University of Illinois Urbana-Champaign)

During reading, skilled readers skip on average between 20-30% of words in text[1]. While studies of adolescents are quite sparse, it is assumed that teens have adult-like behavior[1;2]. In controlled experiments, a combination of low-level visual factors such as word length and higher-level linguistic factors such as word frequency and lexical predictability affect skilled readers' skipping: shorter and lexically more predictable words are skipped more often[3;4]. When it comes to studies investigating these factors in authentic multi-line texts, the results for adult readers are less conclusive[5;cf.6] and have not been examined in adolescent populations. Recently, syntactic predictability (expectations about the upcoming word's part of speech and not the exact word) emerged as an additional source of linguistic predictability that affects skipping in controlled experiments[7;8]. **The present study examined how experience with visual (word length) and linguistic (word frequency, lexical and syntactic predictability) knowledge during reading affects skipping in skilled and adolescent readers in naturalistic texts.**

Method: We recorded eye movements from 113 college students (18-21 yoa; 70 female; 30 male; 3 non-binary) and 52 adolescents (14-17 yoa; 19 female; 30 male; 1 cisgender; 1 gender fluid; 1 nonbinary) who learned English before age 5, while they read 55 passages and total of 2744 words each from the publicly available English language PROVO corpus[9]. Data collection for teens is ongoing. Participants completed vocabulary[10] and reading comprehension (shortened 10-minute version of Nelson & Denny[11]) tests, which were highly correlated with each other for both adults ($r=.50$, $p<.001$) and teens ($r=.59$, $p<.001$) and were combined into one reading experience composite score scaled and centered around grand mean. Human lexical and syntactic predictability scores for each word from the passages are available as part of the PROVO corpus. Word frequencies were obtained from SUBTLEX database[12]. Analyses: A logistic mixed-effects regression model with random intercepts and slopes examined the main effects and the interaction between age, reading experience, word length, word frequency, lexical and syntactic predictability on how many words participants skipped during reading, controlling for the words' sentence position. The study was preregistered. Data will be publicly available upon the completion of this work.

Results: We replicated prior findings from experimental and naturalistic work: readers skip predictable short words more often than predictable long words. Notably, three novel findings emerged (Table 1). First, there was a main effect of reading experience (and not age): better readers skipped more words overall. Second, both age and reading experience interacted with word length but in different ways. Better readers skipped more short and medium words (Figure 2a). Older readers skipped more medium and longer words (Figure 2b). Third, syntactic predictability interacted with word length such that syntactic predictability only affected skipping of medium and longer words (Figure 2c).

Conclusion: Interestingly, words longer than 6 letters were the main source of interactions with reading experience, which may be consistent with deeper parafoveal processing, or other potential explanations. Skipping of short words was uniformly high, but age and linguistic factors affected skipping of medium and long words. Longer words are understudied in eye-tracking studies of reading, pointing to the need for more focused examinations of this portion of the English lexicon where individual and item differences appear. The study also found that adolescents are not quite adult-like in their reading behavior, which calls for more targeted investigations of reading development trajectories to understand the years and decades of reading experience that underlies skilled reading across the lifespan. Finally, this work highlights the role of readers' experience with linguistic knowledge and a word's syntactic predictability in eye movements, suggesting tighter than previously assumed coupling between readers' eye movements and higher-level linguistic predictability beyond specific lexical items.

Table 1. Model Output

Contrast	B	SE	z
Intercept	-.81	.05	-16***
Reading experience	**.16**	**.04**	**3.70*** **
Age	-.04	.04	-1.00
Lexical predictability	**.09**	**.01**	**6.10*** **
Syntactic predictability	.01	.01	1.28
Word frequency	**.21**	**.02**	**11.73*** **
Word length	**-.81**	**.02**	**-41.16*** **
Word position	**.14**	**.01**	**12.65*** **
Reading experience x Age	-.07	.04	-1.70
Reading experience x Lexical predictability	-.01	.01	-1.28
Reading experience x Syntactic predictability	.00	.00	.31
Reading experience x Word frequency	.00	.00	.30
Reading experience x Word length	**.05**	**.01**	**5.11*** **
Age x Lexical predictability	**.02**	**.01**	**3.05** **
Age x Syntactic predictability	-.00	.00	-1.22
Age x Word frequency	**-.02**	**.01**	**-2.86** **
Age x Word length	**.03**	**.01**	**3.50*** **
Word length x Lexical predictability	**.04**	**.02**	**2.16***
Word length x Syntactic predictability	**.03**	**.01**	**2.17***
Word length x Word frequency	**-.18**	**.01**	**-14.29*** **

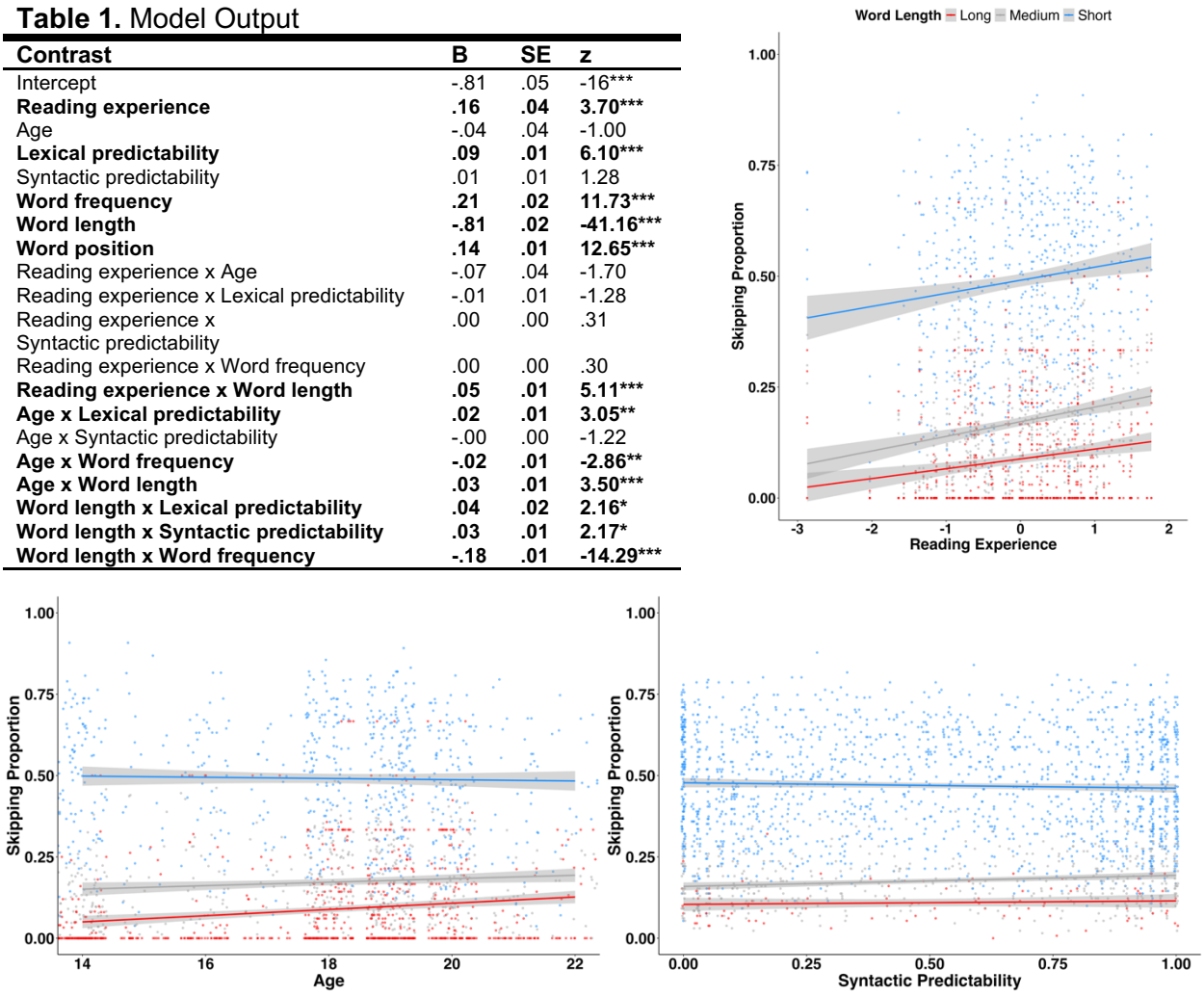

Figure 2a-c. Model estimates of the interactions of word length with the variables of interest are split into 3 groups for visualization purposes only with equal frequency binning (to create comparably equally distributed categories).
Long: 10+ characters; ***Medium***: 6-9 characters; ***Short***: < 5 characters;
Figure 2a & 2b: 1 dot = skipping proportion for 1 participant averaged over all word by length 1-15 letters (total 131 x 15 = 1965);
Figure 2c: 1 dot = skipping proportion for 1 word averaged over all participants (total: 2744)

References

1. Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psych bull*, *124*(3), 372.
2. Rayner, K. (2009). Eye movements in reading: Models and data. *Journal of eye movement research*, *2*(5), 1.
3. Rayner, K., Slattery, T. J., Drieghe, D., & Liversedge, S. P. (2011). Eye movements and word skipping during reading: effects of word length and predictability. *JEP: HPP*, *37*(2), 514.
4. Schotter, E. R., Angele, B., & Rayner, K. (2012). Parafoveal processing in reading. *Attention, Perception, & Psychophysics*, *74*, 5-35.
5. Cevoli, B., Watkins, C., & Rastle, K. (2022). Prediction as a basis for skilled reading. *Royal Society open science*, *9*(6), 211837.
6. Heilbron, M., van Haren, J., Hagoort, P., & de Lange, F. P. (2023). Lexical processing strongly affects reading times but not skipping during natural reading. *Open Mind*, *7*, 757-783.
7. Cutter, M. G., Martin, A. E., & Sturt, P. (2020). The activation of words in syntactically illegal positions. *Qrt JEP*, *73*(9), 1423-1430.
8. Staub, A. (2021). How reliable are individual differences in eye movements in reading?. *JML*, *116*, 104190.
9. Luke, S. G., & Christianson, K. (2018). The Provo Corpus: A large eye-tracking corpus with predictability norms. *BRM*, *50*, 826-833.
10. Shipley, W. C., & Burlingame, C. C. (1941). A convenient self-administering scale for measuring intellectual impairment in psychotics. *American Journal of Psychiatry*, *97*(6), 1313-1325.
11. Gerow, J. R., & Murphy, D. P. (1980). The validity of the Nelson-Denny Reading Test. *Ed and Psych Mnt*, *40*(2), 553-556.
12. Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis. *Behavior research methods*, *41*(4), 977-990.