Studies in Second Language Acquisition (2024), **46**, 1282–1297
doi:10.1017/S0272263124000585

RESEARCH REPORT

Do reading times predict word learning? An eye–tracking study with novel words

Irina Elgort[1] and Elisabeth (Lisi) Beyersmann[2]

[1]Victoria University of Wellington - Te Herenga Waka, Wellington, New Zealand and [2]Macquarie University, Sydney, Australia
Corresponding author: Irina Elgort; Email: irina.elgort@vuw.ac.nz

(Received 29 January 2024; Revised 24 August 2024; Accepted 06 September 2024)

Abstract

Theories of learning and attention predict a positive relationship between reading times on unfamiliar words and their learning; however, empirical findings of contextual learning studies range from a strong positive relationship to no relationship. To test the conjecture that longer reading times may reflect different cognitive and metacognitive processes, the need to infer novel word meanings from context was deliberately manipulated. One hundred and two adult first– and second–language English language speakers read sixty passages containing pseudowords while their eye movements were recorded. The passages were either preceded or followed by pseudoword definitions. After reading, participants completed posttests of cued meaning recall and form recognition. Meaning recall was positively associated with (i) individual cumulative reading times and (ii) participants' general vocabulary knowledge, but not when definitions were provided before reading. Form recognition was unaffected by cumulative reading times. Our findings call for a cautious approach in making causative links between eye–movement measures and vocabulary learning from reading.

Keywords: word learning from reading; eye movements; reading times; vocabulary knowledge

Eye movements in reading reflect ongoing language processing and cognitive processes that underpin it (Rayner, 2009). In this article, we investigate the relationship between the cumulative reading time on novel words *during* reading and the knowledge of these words *after* reading, measured by two posttests of form–meaning mapping: a cued meaning–recall and a form–recognition (gap–fill) posttest. Several eye–movement studies of learning words from reading have found that longer reading times are associated with better word learning (Chaffin, 1997; Chaffin et al., 2001; Godfroid et al., 2013; Williams & Morris, 2004). This aligns with theoretical frameworks that predict that allocating more attention to a novel word during reading leads to better

learning (Robinson, 1995; Schmidt, 1990). However, not all empirical studies have found this positive relationship; some found that longer reading times predicted the learning of meaning but not the learning of form (e.g., Godfroid et al., 2018; Pellicer-Sánchez, 2016; Pellicer-Sánchez et al., 2021), while others did not find any systematic relationship between reading times and vocabulary learning (Elgort et al., 2018; Pellicer-Sánchez et al., 2022). The goal of the present study was to examine whether these inconsistent findings could be explained by variation in the need to infer meanings of unfamiliar words from context. Our findings contribute to a better understanding of the acquisition of novel word meanings from reading in the first and second language.

Reading times and learning gains

As novel words become more familiar with each new encounter in reading, the number and duration of fixations on these words reduce (Elgort et al., 2018; Godfroid et al., 2018; Joseph et al., 2014; Rayner et al., 1995). Pellicer-Sánchez (2016), for example, found that six encounters in a short story were sufficient for first–language (L1) and second–language (L2) readers to start reading six pseudowords (concrete nouns) in the same manner as matched known words, occurring the same number of times in the story (although the learning trajectory was steeper for L1 than L2 readers). However, this reduction in reading times over multiple encounters with a word in reading may result from its increasing familiarity due to its encoding into episodic memory (Reichle & Perfetti, 2003), rather than from its acquisition, a key marker of which is the development of robust form–meaning links (Laufer & Goldstein, 2004; Perfetti, 2007; Rice & Tokowicz, 2020). The development of such links is facilitated by accessing dictionary–style definitions that speed up the encoding of novel words into readers' semantic memory (Bolger et al., 2008; Elgort et al., 2020). (This approach of supplementing contextual word learning with definitions was adopted in the present study.)

The acquisition of L1 and L2 vocabulary from reading has been measured using form and meaning recognition and recall posttests. A positive relationship between attention allocation during reading (measured by total reading time on the learning target) and various posttests (particularly, meaning recall) has been observed in several contextual word learning studies, but these results vary considerably in detail. In an L1 study with pseudowords, Williams and Morris (2004) found that participants' response accuracy on the synonym posttest (measuring the knowledge of meaning) was positively associated with the reading times on a late eye–movement measure (i.e., second pass time) but negatively associated with an earlier eye–movement measure (i.e., gaze duration). In an L2 study, Godfroid et al. (2013) found that more attention (operationalized as longer total reading times) allocated by Dutch students to the pseudowords in English paragraphs was positively associated with their ability to select the right pseudowords for their meaning–bearing context, in a gap–fill posttest of form recognition. Similarly, a positive association between readers' Summed Total Reading Time on the learning targets and their response accuracy on meaning recognition and meaning recall posttests was reported by Godfroid et al. (2018). Pellicer-Sánchez (2016) also found a significant positive relationship between the time spent reading the pseudowords and the accuracy of meaning-recall after reading (but *not* on meaning recognition). Overall, these results suggest that longer reading times on novel words in reading tend to be positively associated with their learning, reflecting the notion that extra time spent during learning reflects deeper processing and richness of encoding

https://doi.org/10.1017/S0272263124000585 Published online by Cambridge University Press

(Robinson, 1995; Schmidt, 1990). However, these findings also highlight the variability of this relationship, observed on some but not other posttests.

An inconsistent relationship between eye movement and posttest measures was recently reported in a study that examined contextual learning of pretaught pseudowords (Pellicer-Sánchez et al., 2021). L1 and L2 readers encountered six pseudowords eight times in a short story, either in a reading–only condition (without vocabulary support) or in a preteaching condition, in which participants memorized pseudowords and their definitions and completed a word–definition matching test before encountering the pseudowords in reading. Participants' knowledge of the pseudowords was assessed in form recognition, meaning recognition, and meaning recall posttests. The total reading time on the first and second occurrences of the pseudowords in the text failed to predict their knowledge on any of the posttests. The cumulative reading time (across eight exposures) was not a significant predictor of learning on meaning-recall or form-recognition, but it predicted response accuracy on the meaning–recognition posttest. Unexpectedly, none of the posttest analyses showed interactions between the learning condition (preteaching versus reading-only) and the cumulative reading times on the pseudowords. This suggests that the advantage of preteaching over reading-only, observed on the posttests, may have been largely due to the learning completed before the reading (in the preteaching procedure) and not the participants' reading behavior on the pseudowords.

The relationship between reading times on novel words and the knowledge of form measured after reading is also mixed. No association between eye–movement measures and accuracy of form recognition or recall was observed in Godfroid et al. (2018) or Pellicer-Sánchez (2016); however, Mohamed (2018) found a positive relationship between both early and late eye–movement measures (namely, first–fixation duration and total reading times) and form recognition after reading. Similarly, a positive relationship between reading times (the total reading time and second pass reading time) on unknown French words in the captions and their learning was reported by Montero Perez et al. (2015) in contextual word learning from video input (clips from a Swiss and Belgian current affairs program). However, this positive relationship was only observed in the intentional learning condition, when participants were informed about an upcoming vocabulary posttest. When participants watched clips for meaning and were not aware of the vocabulary learning goal, longer reading times were negatively associated with learning (measured by a form recognition posttest). Montero Perez et al. reported that this negative relationship was observed on the second pass reading time – a late eye–movement measure which, in the incidental learning condition, likely reflected processing effort (such as word–to–text integration) rather than an intention to acquire a word.

Eye movements in reading unfamiliar words

One reason for these mixed findings is that readers' eye movements on novel and unfamiliar words in natural reading may be affected by a host of variables, such as reading goals, the use of cognitive and metacognitive strategies, individual differences, and text difficulty (Godfroid et al., 2018, 2020; Wang & Pellicer-Sánchez, 2023). Understanding how these factors modulate the relationship between reading times and vocabulary learning is useful in building a more detailed and realistic picture of contextual vocabulary learning – a key source of increasing vocabulary size in the first (L1) or second (L2) language (e.g., Beck et al., 2002; Grabe, 2009; Nation, 2006, 2022). In

https://doi.org/10.1017/S0272263124000585 Published online by Cambridge University Press

addressing the question of whether longer reading times predict better word learning outcomes, we turned to two key reasons for the slower reading of novel words: processing effort (reflecting difficulty) and learning effort (reflecting engagement, depth of processing, and richness of encoding, e.g., Craik & Lockhart, 1972; Craik & Tulving, 1975).

Readers make more and longer fixations when they experience processing difficulty; for example, when texts are more challenging or when readers are less skilled (Rayner, 2009). Lower–frequency words are more difficult to process than higher–frequency words (Brysbaert et al., 2017); eye–movement studies consistently show that lower–frequency words are read slower than higher–frequency words (Inhoff & Rayner, 1986; Just & Carpenter, 1980; Rayner & Duffy, 1986; Juhasz & Rayner, 2006; Kliegl et al., 2004), in L1 and L2 (Cop et al., 2015). Because language users are exposed to high–frequency words more often, their lexical representations are more accessible in real–time processing, such as reading, and require less processing and attentional resources. Words that have not been encountered, on the other hand, have zero frequency and remain low frequency in the early stages of learning (Elgort et al., 2018; Godfroid et al., 2018). They do not have lexical representations that can be automatically activated and draw more processing and attentional resources during reading. To recap, novel words are read more slowly and with more effort because they are *difficult to process*.

Eye movements in reading also reflect the *learning effort* associated with reading unfamiliar words. Personal experiences of individual readers with the word affect the number and duration of fixations, particularly, on lower–frequency words (Chaffin et al., 2001; Juhasz, 2022; Juhasz & Pollatsek, 2011; Williams & Morris, 2004). Longer reading times on unfamiliar than familiar words are interpreted as readers' effort to learn them, that is, orthographically encode their forms, infer meanings, and integrate these meanings into context (Chaffin, 1997; Chaffin et al., 2001; Lowell & Morris, 2014). This speaks to the second possible cause of slower reading times: unfamiliar words are read more slowly because readers make an effort to learn them.

The hypotheses that longer reading times reflect processing difficulty on the one hand and learning effort on the other are not necessarily mutually exclusive. Instead, longer reading times on novel words likely reflects a complex interplay between processing difficulties and learning effort, which may explain inconsistent findings regarding whether longer reading times on novel words during reading predict how well they are known after reading. The present study investigated this conjecture by using a novel word learning paradigm that either encouraged readers to infer novel word meanings during reading (increasing their learning effort) or reduced the need for it[1].

Present study

In the present study, we manipulated the need to infer novel word meanings from context by giving participants access to definitions of novel words (pseudowords) either

[1]Note that, although the data used in the present article are the same as in Elgort et al. (2023), the present analysis addresses a fundamentally different research question. The goal of Elgort et al. (2023) was to tests whether previewing novel words and their definitions *before* reading changed L1 and L2 readers' eye-movement patterns on these words *during* reading. A one-page Accessible Summary of Elgort et al. (2023) is available at https://oasis-database.org. The present study aims to contribute to a more nuanced understanding of the relationship between eye movements *during* reading and the outcomes of contextual word learning, measured by meaning recall and form recognition.

https://doi.org/10.1017/S0272263124000585 Published online by Cambridge University Press

before (PRE) or after (POST) encountering them in reading. In both conditions, participants were instructed to "explain the meaning of each pseudoword, as understood from the text" immediately after reading. The purpose of explaining the meaning was to create a need for contextual learning without explicitly instructing participants to learn the pseudowords. The resulting participants' knowledge of the pseudowords was measured using cued meaning–recall and form–recognition posttests. The main research question addressed in this study was whether the relationship between reading times on unfamiliar words and their learning is modulated by the degree of need to derive their meanings from context during reading (operationalized as the PRE– and POST–condition).

We predicted that when definitions of the pseudowords were not provided prior to reading (POST–condition), longer reading times on the pseudowords would be more likely to reflect the learning effort (e.g., participants' attempts to encode novel word forms and infer their meanings in reading, and to integrate their meanings into context). In the PRE–condition, on the other hand, reading times were predicted to be less affected by the learning effort *during* reading, since participants had been exposed to the pseudowords and their meanings *before* reading. In fact, longer reading times in the PRE–condition may index processing difficulty (e.g., due to an incomplete understanding of the meaning presented in the definition, difficulties in reconciling the meanings in the preview with the meanings in the text, or challenges in form–meaning mapping). This is not to say that some learning effort and processing difficulties are not present in both conditions, to some extent. However, due to the differences in the need to infer novel word meanings from context, the cumulative reading times (CRT) on the pseudowords would likely predict participants' performance on the posttests in the POST–condition but not in the PRE–condition.

Our secondary research question was whether a predicted relationship between CRT and learning outcomes was modulated by the participants' linguistic differences at an individual and group level. Participants' lexical knowledge and language group (L1 vs. L2) have been shown to affect reading times (Kuperman & Van Dyke, 2011; Kuperman et al., 2023) and contextual word learning (Elgort et al., 2015; Elgort & Warren, 2014; Frishkoff et al., 2008). To this end, the present study was conducted with adult L1 and L2 speakers of English, who varied in their English lexical proficiency. Two measures of individual differences were included in the data analysis: participants' vocabulary knowledge – a continuous measure based on the Lexical Test for Advanced Learners of English (LexTALE; Lemhöfer & Broersma, 2012) and their language group (L1/L2). We predicted that inferring word meanings from context would be more challenging for participants with less vocabulary knowledge, and these participants might be less successful in establishing accurate form–meaning mappings (Hu & Nation, 2000; Perfetti, 2007; Stanovich, 1986). Therefore, participants with lower LexTALE scores were predicted to score lower on the posttests, compared with participants with higher LexTALE scores. We also conjectured that the impact of the learning condition might be modulated by participants' lexical proficiency and language group; namely, in the POST–condition, higher general reading fluency of L1 readers and readers with larger vocabularies may free up cognitive resources for both orthographic and semantic encoding of pseudowords during reading. These individual differences were predicted to matter less in contextual learning when pseudowords and their definitions were previewed before reading. This is because previewing pseudowords and their definitions before reading reduces the burden of contextual inferencing during reading, which may specifically support L2 readers and those with smaller vocabularies by increasing their ability to

https://doi.org/10.1017/S0272263124000585 Published online by Cambridge University Press

Table 1. Participants' details. Averages and standard deviations (in parenthesis) by language group

	L1 participants	L2 participants
Age (in years)	21 (8.6)	27 (6.0)
LexTALE	86% (9.1)	73% (13.8)
Equivalent CERF level	C1/C2 (advanced)	B2 (upper intermediate)
TOWRE–2		
Raw score / word sight efficiency	96 (8.2) / average performance	86 (12.3) / slightly below the average range
Raw score / phonemic decoding efficiency	58 (6.8) / average performance	54 (10.0) / average performance

Note. CERF - Common European Framework of Reference for Languages; TOWRE-2 - Test of Word Reading Efficiency, 2nd edition, Torgesen et al. (2012).

encode pseudowords into memory while reading, thereby reducing the differences in learning outcomes between the L1 and L2 participant groups.

In summary, we analyzed response accuracy on the cued meaning–recall and form–recognition posttests of 60 contextually learned pseudowords as a function of (1) individual cumulative reading times on the pseudowords and (2) learning condition, to test whether the relationship between reading times and learning outcomes was modulated by the level of need to infer word meanings from context, constrained by the learning condition. We also checked whether participants' vocabulary knowledge and their language group (L1/L2) affected their posttest scores differently in the two learning conditions.

Method

Participants

One hundred and two[2] L1 (n = 49) and L2 (n = 53) adults participated in the study[3] (Table 1). L1 participants were undergraduate psychology students who participated in the study to fulfill course requirements. L2 participants were undergraduate or graduate university students, who received a modest incentive for their participation. Their L1 included Chinese (15), Vietnamese (7), Tamil (4), Japanese (3), and two participants each from French, Hindi, Italian, Malayalam, and Tagalog (over 90% of the participants' L1s had non–Latin scripts). The English language proficiency of L2 participants was at level 6.0 or higher on the International English Language Testing System, aligned with the university's admission requirements.

[2]An a priori power analysis was based on Elgort et al. (2020) - the first study, to our knowledge, that tested the effect of definitions' placement (before and after reading) on contextual word learning. A power simulation conducted using simr (Green & MacLeod, 2016) showed that 100% (95% CIs: 96.38, 100.00) power could be achieved for the main effect of the learning condition with 60 participants and 60 items, and 80% power with 25 participants. Therefore, 100 participants (and 60 items) are predicted to provide sufficient power for the present analysis that tested a two-way interaction (Condition × CRT).

We also conducted a post hoc power simulation for the predictor Condition, in the MG posttest accuracy data model, with a two-way interaction between (Condition × CRT). We found that this analysis was sufficiently powered for the Condition predictor (power = 96.00%, 95% CIs: 90.07, 98.90).

[3]Initially, data from 50 L1 participants and 57 L2 participants were collected, but data from four L2 participants were excluded due to poor calibration and one L1 participant who did not follow instructions.

https://doi.org/10.1017/S0272263124000585 Published online by Cambridge University Press

Materials

A complete description of the materials and procedure is provided in Elgort et al. (2023) and online (see https://osf.io/2um7b). Here, we provide only key details needed to interpret the present study. Participants read 60 short passages related to the topics of building/housing and medicine/health, averaging 116 words per passage. Passages contained one novel item (pseudoword) repeated three times within each passage. Orthographically and phonologically legal English pseudowords, six to seven letters long (e.g., "frount") were selected from the MCWord database (Medler & Binder, 2005) as the learning targets. To support readers' contextual inferences, a word or phrase closely related in meaning to the pseudoword was included in the passage, before the first occurrence of the pseudoword (e.g., "Different injections may require the use of a different ruggle and needles."; "Welding works well for joining metals with similar melting points but triting is used to join dissimilar metals by choosing a filler metal with a lower melting point than either of the metals to be joined.").

Short dictionary–type definitions of the pseudowords were, on average, 14 words long (SD = 4.5). In both definitions and reading passages, 98% of the running words were within the 6,000 most frequent word families of English, making the reading materials appropriate for the L2 participants in the present study (Drummond, 2018). The critical manipulation was whether definitions of the pseudowords were presented before (PRE) or after (POST) reading. In a counterbalanced experimental design, half of the pseudowords were presented in each of the two learning conditions that were intermixed (not blocked). In the POST–condition, the text was the only source of information about pseudoword meanings, at the time of reading. In contrast, in the PRE–condition, participants previewed pseudowords and their definitions before reading and were therefore aware of the novel word meanings during the reading task.

Posttests

Outcomes of contextual word learning were measured using a meaning generation (MG) posttest and a multiple–choice (MC) form–recognition posttest. In the MG posttest, pseudowords were presented in short neutral sentences (e.g., "It takes time to frount as you know") and participants were asked to define their meaning; in the MC posttest, participants chose from four word-forms (pseudowords) to complete a sentence that constrained the pseudoword's meaning (e.g., "The doctor, puzzled by my symptoms, could not _____ my condition" [answer: "frount"; similar in meaning to "diagnose"]). Thus, in the MG test, participants had to recall the pseudoword's meaning when its form was provided (form → meaning), without relying on contextual support and, in the MC test, pseudoword meanings could be inferred from the sentence and participants had to select the pseudoword to match that meaning (meaning → form). Although both posttests measured form–meaning mapping, the MG posttest emphasized meaning recall, and the MC posttest form recognition.

Procedure

The data were collected in two lab sessions, on two days (participants read 30 passages per day). Each pseudoword trial consisted of the following components presented on separate screens (Figure 1): (1) the reading passage containing three instances of the pseudoword, (2) the pseudoword with its definition (presented before the passage screen in the PRE–condition and after the passage screen in the POST–condition), and

https://doi.org/10.1017/S0272263124000585 Published online by Cambridge University Press

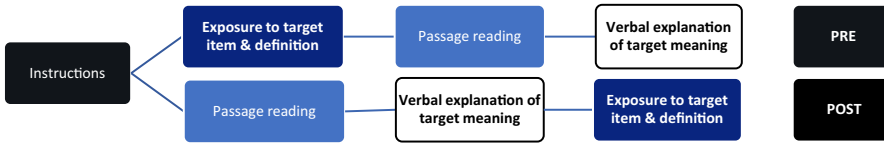

Figure 1. Diagrammatic representation of the two learning conditions (Elgort et al., 2023).

(3) a screen where the pseudoword was presented in isolation, and participants verbally explained its meaning, as inferred from the text (always presented immediately after the passage screen). Participants were not explicitly instructed to learn the pseudowords, but they knew they would need to explain the meanings of the pseudowords after reading. This created an explicit need to map novel word forms onto meanings. After each learning session, participants completed an intervening task, the MG posttest, and the MC posttest, in that order.

Eye movements were recorded using an EyeLink 1000 (SR Research Ltd., Canada) eye-tracker (sampling rate of 1,000 Hz). Each passage was presented on a single screen, using black Courier New font size 20, on a light grey background, the lines were triple-spaced, and the pseudoword never appeared at the start or end of a line. A nine–point calibration was executed before the beginning of the reading procedure and after each break. A drift check was performed before each reading passage; additional calibrations were performed for each participant, as required.

Analysis and results

The MG and MC posttests were analyzed separately. We modeled the posttest response accuracy as a function of the learning condition (Condition), CRT on the pseudowords, individual differences, and their interactions. In choosing CRT as a measure of attention allocated to the pseudowords during reading we were motivated by the fact that this measure is most strongly correlated with contextual vocabulary learning (Godfroid, 2020; see also Godfroid et al., 2018, p. 6).

The glmer models (lme4 package in R, Bates et al., 2015) were fitted to the binary accuracy data (1 = correct/0 = incorrect), with participants and items as crossed–random effects. CRTs were calculated as a sum of the total reading times on the three occurrences of the pseudoword in the text, for each participant. After removing extreme outliers shorter than 300 and longer than 7,000 milliseconds (1.4%), CRT was log-transformed and normalized, using the scale() function in R. Condition (PRE/POST) and Language (L1/L2) were contrast coded (-.05/0.5, respectively). The LexTALE scores were residualized over Language (r = 0.87), to avoid multicollinearity, and centered. The initial models, in both analyses, included a three–way interaction between CRT, Condition, and LexTALE, and a two–way interaction between Condition and Language. The final models were identified following the backward stepwise model simplification procedure, using the likelihood ratio test for model comparisons (Baayen et al., 2008). We attempted to fit the maximal random–effects structure justified by the data, unless convergence issues were encountered (Barr et al., 2013; Matuschek et al., 2017). The emmeans package (Lenth, 2019) was used to conduct post hoc analyses, with Bonferroni p-value correction for multiple comparisons. Analysis of deviance (using type II Wald chi-square tests) was used as a test of effects.

The results of the MG and MC posttests were reliable (Cronbach's alpha α = .91 and α = .92, respectively). The mean MG accuracy was 0.29 (95% CI: 0.28, 0.30); the mean

https://doi.org/10.1017/S0272263124000585 Published online by Cambridge University Press

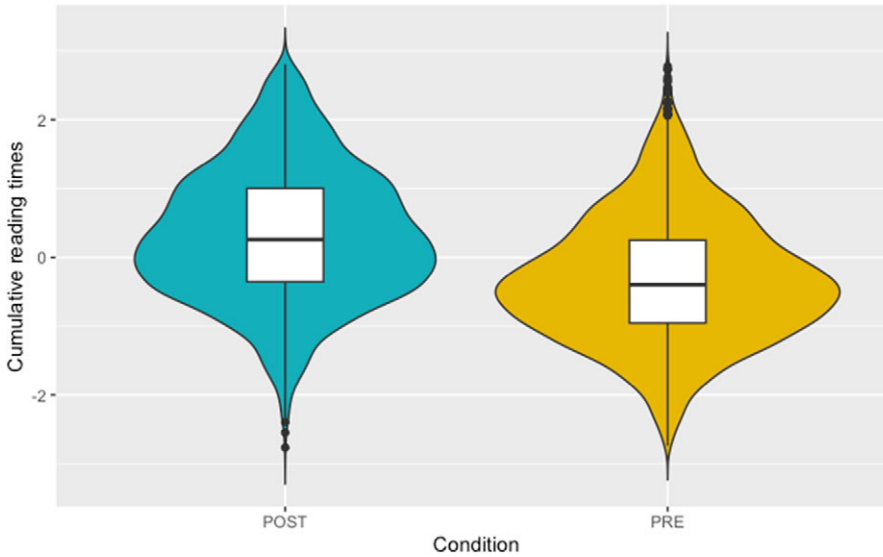

Figure 2. Cumulative reading times by condition (based on descriptive statistics).

MC accuracy was 0.70 (95% CI: 0.69, 0.71). See Table 2 for descriptive statistics. Participants' CRT (Figure 2) was longer in the POST– than in the PRE–condition ($F_{(1,109)} = 359.83$ $p < .001$).

Meaning–generation posttest

In the analysis of deviance (Appendix A, Tables 1.1–1.3), two statistically significant interactions were confirmed (Figure 3): between Condition and CRT ($\chi^2 = 10.59$, $p = .001$) and between Condition and participants' vocabulary knowledge ($\chi^2 = 6.59$, $p = .010$). Post hoc comparisons showed that, in the POST–condition, longer CRT resulted in significantly more accurate meaning recall than shorter CRT ($z = -2.36$, $p = .019$); conversely, in the PRE–condition, longer CRT resulted in less accurate meaning recall than shorter CRT, but this contrast did not reach statistical significance ($z = 1.66$, $p = .098$). In other words, in the POST–condition, there was a clear positive relationship between CRT and the learning of meaning but, in the PRE–condition, we

Table 2. Descriptive statistics by learning condition and language group

		MG			MC			CRT		
			95% CIs			95% CIs			95% CIs	
Language	Condition	Mean	Lower	Upper	Mean	Lower	Upper	Mean	Lower	Upper
L1	POST	0.25	0.23	0.27	0.73	0.71	0.76	1450	1416	1484
L1	PRE	0.26	0.23	0.28	0.70	0.68	0.73	1046	1024	1069
L2	POST	0.33	0.30	0.35	0.68	0.65	0.70	2509	2446	2573
L2	PRE	0.32	0.29	0.34	0.70	0.67	0.72	1678	1632	1725

https://doi.org/10.1017/S0272263124000585 Published online by Cambridge University Press

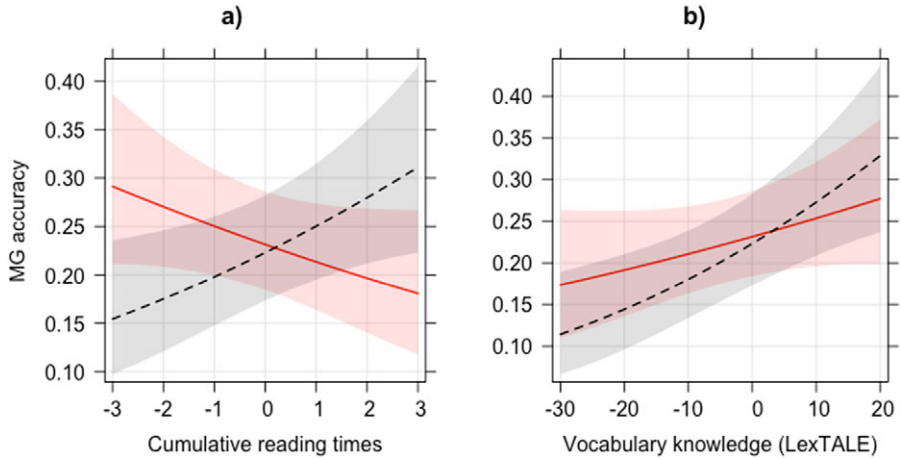

Figure 3. Meaning–generation posttest. Partial interaction plots: (a) Condition × Cumulative Reading Time and (b) Condition × LexTALE.
Note. Dashed black lines indicate POST–condition and solid red lines indicate PRE–condition.

observed an opposite trend; namely, longer reading times were associated with weaker meaning recall.

LexTALE was a significant predictor of learning in the POST–condition (z = -3.13, p < .01), but not in the PRE–condition (z = -1.50, p = .133), showing that meaning recall in the POST–condition was greater in participants with larger vocabulary knowledge. Language did not predict meaning recall and did not interact with Condition.

Multiple–choice cued form–recognition posttest

CRT did not predict form–recognition accuracy (χ^2 = 1.88, p = .170) and did not interact with Condition or LexTALE. There were significant main effects of Condition (POST > PRE: χ^2 = 4.42, p = .035) and participants' vocabulary knowledge (χ^2 = 29.16, p < .001), showing better pseudoword recognition accuracy in the POST–condition and for participants with higher LexTALE scores (Appendix B, Tables 2.1-2.2). There was no main effect of Language, but an interaction between Condition and Language improved the model fit (χ^2 = 4.55, p = .033). The post hoc analysis showed that L1 participants were more accurate in the POST– than PRE–condition (z = -2.10, p = .035), but this was not the case for L2 participants (z = 0.75, p = .452).

Discussion

To investigate whether reading times on initial contextual encounters with novel words predict their learning and address inconsistent findings about this relationship in the literature, we manipulated learning conditions to either increase the need to infer pseudoword meanings during reading or to reduce it. In the POST–condition, the need to infer pseudoword meanings from context was increased because the reading text was the only source of information before participants had to explain the meanings of the novel items. In the PRE–condition, the need to infer pseudoword meanings from context was reduced because pseudowords and their definitions had been previewed

https://doi.org/10.1017/S0272263124000585 Published online by Cambridge University Press

before reading, and participants could use this information in explaining the meanings of the items. We recorded L1 and L2 participants' eye movements while they were reading the passages containing pseudowords and estimated their learning using meaning–recall and form–recognition posttests.

In line with our hypotheses, in the POST–condition (high contextual word learning effort), longer CRT predicted better knowledge of meaning, confirming a positive relationship between attention to unfamiliar words during reading and their meaning recall (Robinson, 1995; Schmidt, 1990). One possible explanation for this finding is that engaging in contextual lexical inferencing involves cognitive effort that is associated with learning and retention (Bjork & Kroll, 2015; Elgort et al., 2020). Information about novel word meanings generated in the process of contextual inferencing likely results in robust memory traces (Bertsch et al., 2007). Therefore, it is reasonable to argue that longer reading times in conditions that encourage contextual meaning inferences are likely to be positively associated with the learning of meaning. This is also consistent with the more general notion that more fixations and longer fixations on a target stimulus during learning allow for more effective encoding of information into long–term memory (Damiano & Walther, 2019; Kafkas & Montaldi, 2011). However, when the need to infer word meanings was reduced by providing pseudoword definitions before reading, longer reading times were no longer positively associated with better knowledge of meaning; instead, they were associated with somewhat *less* accurate meaning recall (although this contrast did not reach statistical significance in the post hoc analysis). Note that, on average, participants spent less time on the novel items in the PRE– than in the POST–condition (Figure 2), confirming a reduced perceived need to attend to the pseudowords during reading in the PRE–condition.

This finding has important implications for vocabulary acquisition studies that use eye-tracking as a proxy for contextual word learning. We have shown that CRTs on novel words in reading do not only index learning; in other words, it cannot be assumed that the relationship between attention allocation in natural reading and vocabulary learning from reading should always be positive. Lawson and Hogben (1996), for example, point out an important distinction between inferring word meanings for text comprehension and learning. For readers who focus on engaging with the ideas expressed in the text, longer reading times on novel words may index difficulties in language processing and/or comprehension and are less likely to be positively associated with vocabulary learning. On the other hand, when readers are motivated to increase their vocabulary knowledge and use vocabulary learning strategies, they tend to be more successful in contextual word learning (Elgort & Warren, 2014), and their CRTs on novel words are more likely to predict learning. These individual differences in how readers approach novel words are more likely to affect the relationship between attention allocation in reading and word learning in studies with longer authentic texts (Elgort et al., 2018; Godfroid et al., 2018) than in studies with shorter texts with highly controlled lexical frequency profiles. This is because, in short, simplified texts, even weaker readers may afford to allocate cognitive resources to overt lower–order processes, including mapping of novel word forms onto contextually appropriate meanings, without compromising higher–order reading comprehension processes (e.g., Graesser et al., 1997). This may be why contextual word–learning studies that use short texts (Godfroid et al., 2013; Pellicer-Sánchez, 2016) tend to report a positive relationship between CRTs and word learning.

A second key finding was that, when participants inferred pseudoword meanings during reading before seeing their definitions (POST–condition), having higher vocabulary knowledge led to more accurate meaning recall, but this was not the case when

https://doi.org/10.1017/S0272263124000585 Published online by Cambridge University Press

definitions were previewed before reading in the PRE–condition. The learning advantage, where better readers with larger vocabularies are better able to learn words from reading, known as the Matthew effect (Stanovich, 1986) is well documented in the L1 and L2 literature. Readers with larger vocabularies achieve better text comprehension (Cromley & Azevedo, 2007; Jeon & Yamashita, 2022; Rott, 1999) and can better use contextual clues to derive unfamiliar word meanings (e.g., Schmitt et al., 2011), which improves their chances of acquiring new vocabulary from unassisted reading (Elgort & Warren, 2014). Extensive vocabulary knowledge is also associated with denser lexical–semantic networks that amplify resonance processes, by which correctly inferred semantic features of novel words resonate with overlapping clusters of features in learners' existing lexical–semantic networks (Rodd, 2020, Perfetti, 2007; Reichle & Perfetti, 2003), facilitating the encoding of new knowledge into semantic memory. In the PRE–condition, on the other hand, participants were exposed to the core meaning features of the pseudowords upfront, via definitions (Bolger et al., 2008), which modulated the learning advantage associated with having a larger vocabulary.

Finally, like most previous contextual word learning studies (Godfroid et al., 2018; Pellicer-Sánchez, 2016; Pellicer-Sánchez et al., 2021), we did not find a reliable relationship between CRT and form–recognition accuracy in either learning condition, suggesting that longer reading times on novel words are not necessarily beneficial for developing formal–lexical representations (cf. Elgort et al., 2016, who found that actively deriving novel word meanings during L2 reading was less beneficial for developing their precise lexical representations than practicing their spelling). The finding that L1 (but not L2) readers were better at form recognition in the POST– than PRE–condition supports the L1 reading advantage hypothesis (namely, higher reading efficiency in L1 affords more precise orthographic encoding of novel words during reading, even when meaning inferences are prioritized by the reader).

Limitations and directions for future research

Eye movements on novel words in reading are affected by a combination of factors, including individual readers' tasks and reading behaviors that may not have been fully accounted for by using mixed–effects regressions. However, against the background of the overall shorter reading times on the pseudowords in the PRE– than the POST–condition (as shown in Elgort et al., 2023), the negative trend in the relationship between longer total reading times and knowledge of meaning gains suggests that additional attention (operationalized as longer CRTs) was not useful for learning in the PRE–condition. A promising avenue for future research would be triangulating eye–movement data and posttest measures with participants' own account of attention allocation and vocabulary learning during reading in the two conditions, to address the question of why positive effects of longer reading times were not observed in the PRE–condition (Godfroid & Schmidtke, 2013; Wang & Pellicer-Sánchez, 2023).

A further limitation is that word learning was examined with explicit vocabulary support, in the form of definitions. Moreover, although participants were not instructed to learn pseudowords encountered in reading, the experimental procedure required them to explain their meanings after reading. These design choices created learning conditions that are different from purely incidental reading behavior. As such, individual readers' word learning and inferencing behaviors may have differed considerably as a result of their applied reading goals and language learning strategies (Elgort & Warren, 2014; Lawson & Hogben, 1996; Pulido, 2007). Future exploratory reading–

https://doi.org/10.1017/S0272263124000585 Published online by Cambridge University Press

only studies may clarify how individual approaches to reading unfamiliar words affect the relationship between eye movements and learning gains, under more incidental learning conditions.

In this research report, we tested whether the overall (cumulative) reading time readers engage with novel words in a single learning event (reading a paragraph) is necessarily predictive of learning. Future studies may choose to approach this issue from a different perspective; for instance, researchers may ask what component eye movements (e.g., skipping, first–fixation duration, gaze duration, total reading time, regressions) on novel words in reading are associated with word learning, or whether trajectories of change in the total reading time across contextual encounters are predictive of novel word learning from reading.

Conclusion

Recording eye movements using modern eye trackers and software has revolutionized the study of reading and created opportunities for gaining new insights into contextual word learning in real-time (Joseph et al., 2014; Godfroid et al., 2013). However, reading and language learning researchers should heed the warning against making uncritical connections between eye–movement data and "what precisely went on in the reader's mind" (Boers, 2022, p. 13). In this study, we were able to tease apart one reason why a theoretically motivated positive relationship between cumulative reading times on initial encounters with unfamiliar words *during* reading and the knowledge of these words measured *after* reading is not always observed. Since engaging additional attentional resources for contextual word learning and dealing with processing difficulties in reading both result in longer total reading times, making strong directional predictions about word learning outcomes based on attention allocation during reading may be misguided, unless the study design, instructions, or readers' intentions prioritize word learning.

Supplementary material. The supplementary material for this article can be found at http://doi.org/10.1017/S0272263124000585.

Competing interest. The authors declare none.

References

Baayen, R., Davidson, D., & Bates, D. (2008). Mixed-effects modelling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412.

Barr, D.J., Levy, R., Scheepers, C., & Tily, H.J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255–278.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48.

Beck, I. L., McKeown, M. G., & Kucan, L. (2002). *Bringing words to life: Robust vocabulary instruction*. Guildford Press.

Bertsch, S., Pesta, B. J., Wiscott, R., & McDaniel, M. A. (2007). The generation effect: A metaanalytic review. *Memory & Cognition*, *35*(2), 201–210.

Bjork, R. A., & Kroll, J. F. (2015). Desirable difficulties in vocabulary learning. *The American Journal of Psychology*, *128*(2), 241–252.

Boers, F. (2022). Glossing and vocabulary learning. *Language Teaching*, *55*(1), 1–23.

Bolger, D.J., Balass, M., Landen, E. & Perfetti, C. A. (2008). Contextual variation and definitions in learning the meaning of words. *Discourse Processes*, *45*, 122–159.

Brysbaert, M., Lagrou, E., & Stevens, M. (2017). Visual word recognition in a second language: A test of the lexical entrenchment hypothesis with lexical decision times. *Bilingualism: Language and Cognition*, *20*, 530–548.

Chaffin, R. (1997). Associations to unfamiliar words: Learning the meaning of new words. *Memory and Cognition*, *25*, 203–226.

Chaffin, R., Morris, R.K., & Seely, R.E. (2001). Learning new word meanings from context: A study of eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 225–235.

Cop, U., Keuleers, E., Drieghe, D., & Duyck, W. (2015). Frequency effects in monolingual and bilingual natural reading. *Psychonomic Bulletin & Review*, *22*, 1216–1234.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671–684.

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, *104*(3), 268–294.

Cromley, J. G., & Azevedo, R. (2007). Testing and refining the direct and inferential mediation model of reading comprehension. *Journal of Educational Psychology*, *99*(2), 311–325.

Damiano, C., & Walther, D. B. (2019). Distinct roles of eye movements during memory encoding and retrieval. *Cognition*, *184*, 119–129.

Drummond, A. (2018). Investigating the relationship between IELTS scores and receptive vocabulary size. *Journal of the Foundation Year Network*, *1*, 113–125.

Elgort, I., Beliaeva, N., & Boers, F. (2020). Contextual word learning in the first and second language: Definition placement and inference error effects on declarative and nondeclarative knowledge. *SSLA*, *42*(1), 7–32.

Elgort, I., Brysbaert, M., Stevens, M., & Van Assche, E. (2018). Contextual word learning during reading in a second language: An eye-movement study. *Studies in Second Language Acquisition*, *40*, 341–366.

Elgort, I., Candry, S., Eyckmans, J., Boutorwick, T.J., & Brysbaert, M. (2016). Contextual word learning with form-focused and meaning-focused elaboration. *Applied Linguistics*, *39*(5), 646–667.

Elgort, I., Perfetti, C.A., Rickles, B., & Stafura, J.Z. (2015). Contextual learning of L2 word meanings: Second language proficiency modulates behavioural and ERP indicators of learning. *Language, Cognition and Neuroscience*, *30*(5), 506–528.

Elgort, I., & Warren, P. (2014). L2 vocabulary learning from reading: Explicit and tacit lexical knowledge and the role of learner and item variables. *Language Learning*, *64*, 365–414.

Elgort, I., van de Wetering, R., Arrow, T., & Beyersmann, E. (2023). Previewing Novel Words Before Reading Affects Their Processing During Reading: An Eye-Movement Study With First and Second Language Readers. *Language Learning*.

Frishkoff, G. A., Collins-Thompson, K., Perfetti, C. A., & Callan, J. (2008). Measuring incremental changes in word knowledge: Experimental validation and implications for learning and assessment. *Behavior Research Methods*, *40*, 907–925.

Godfroid, A. (2020). *Eye tracking in second language acquisition and bilingualism: A research synthesis and methodological guide*. Routledge.

Godfroid, A., Ahn, J., Cho, I., Ballard, L., Cui, Y., Johnston, S., Lee, S., Sarkar, A., & Yoo, H.J. (2018). Incidental vocabulary learning in a natural reading context: an eye-tracking study. *Bilingualism: Language and Cognition*, *21*(3), 563–584.

Godfroid, A., Boers, F., & Housen, A. (2013). An eye for words: Gauging the role of attention in incidental L2 vocabulary acquisition by means of eye tracking. *Studies in Second Language Acquisition*, *35*, 483–517.

Godfroid, A., & Schmidtke, J. (2013). What do eye movements tell us about awareness? A triangulation of eye-movement data, verbal reports and vocabulary learning scores. In J.M. Bergsleithner, S.N. Frota, & J.K. Yoshioka (Eds.), *Noticing and second language acquisition: Studies in honor of Richard Schmidt* (pp. 183–205). University of Hawai'i, National Foreign Language Resource Center.

Grabe, W. (2009). *Reading in a second language*. Cambridge University Press.

Graesser, A. C., Millis, K. K., & Zwaan, R. A. (1997). Discourse comprehension. *Annual Review of Psychology*, *48*, 163–189.

Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493–498.

Hu, M., & Nation, I.S.P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language*, *13*, 403–430.

Inhoff, A. W., & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception and Psychophysics*, *40*(6), 431–439.

Jeon, E. H., & Yamashita, J. (2022). L2 reading comprehension and its correlates: An updated meta-analysis. In: Jeon, E.H. & In'nami, Y. *Understanding L2 Proficiency: Theoretical and meta-analytic investigations*. John Benjamins. pp. 29—86.

Joseph, H. S. S. L., Wonnacott, E., Forbes, P., & Nation, K. (2014). Becoming a written word: Eye movements reveal order of acquisition effects following incidental exposure to new words during silent reading. *Cognition*, *133*, 238–248.

Juhasz, B. J. (2022). Using eye movements to investigate the impact of childhood and recent frequency of occurrence on word identification during reading in college. *Scientific Studies Of Reading*, *26*(5), 409–416.

Juhasz, B. J., & Pollatsek, A. (2011). Lexical influences on eye movements in reading. In S. P. Liversedge, I. Gilchrist & S. Everling (Eds.), *The Oxford Handbook of Eye Movements* (pp. 874–893). Oxford University Press.

Juhasz, B.J., & Rayner, K. (2006). The role of age of acquisition and word frequency in reading: Evidence from eye fixation durations. *Visual Cognition*, *13*, 846–863.

Just, M., & Carpenter, P. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, *87*(4), 329–354.

Kafkas, A., & Montaldi, D. (2011). Recognition memory strength is predicted by pupillary responses at encoding while fixation patterns distinguish recollection from familiarity. *Quarterly Journal of Experimental Psychology*, *64*(10), 1971–1989.

Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology*, *16*, 262–284.

Kuperman, V., & Van Dyke, J. A. (2011). Effects of individual differences in verbal skills on eye-movement patterns during sentence reading. *Journal of Memory and Language*, *65*(1), 42–73.

Kuperman, V., Siegelman, N., … Usal, K. A. (2023). Text reading in English as a second language: Evidence from the Multilingual Eye-Movements Corpus. *Studies in Second Language Acquisition*, *45*(1), 3–37.

Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning*, *54*, 399–436.

Lenth, R. (2019). emmeans: Estimated Marginal Means, aka Least-Squares Means. *R package version 1.4.1.*

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavioral Research Methods*, *44*, 325–43.

Lowell, R., & Morris, R.K. (2014). Word length effects on novel words: Evidence from eye movements. *Attention, Perception, & Psychophysics*, *76*, 179–189.

Lawson, M. J., & Hogben, D. (1996). The vocabulary-learning strategies of foreign-language students. *Language Learning*, *46*, 101–135.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315.

Medler, D.A., & Binder, J.R. (2005). *MCWord: An on-line orthographic database of English language.* Retrieved from http://www.neuro.mcw.edu/mcword/

Mohamed, A. (2018). Exposure frequency in L2 reading: An eye-movement perspective of incidental vocabulary learning. *Studies in Second Language Acquisition*, *40*(2), 269–293.

Montero Perez, M., Peters, E., & Desmet, P. (2015). Enhancing vocabulary learning through captioned video: An eye-tracking study. *The Modern Language Journal*, *99*(2), 308–328.

Nation, I. S. P. (2022). *Learning vocabulary in another language*. Cambridge University Press.

Nation, P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review*, *63*, 59–82.

Pellicer-Sánchez, A. (2016). Incidental L2 vocabulary acquisition from and while reading: An eye-tracking study. *Studies in Second Language Acquisition*, *38*, 97–130.

Pellicer-Sánchez, A., Conklin, K., & Vilkaitė-Lozdienė, L. (2021). The effect of prereading instruction on vocabulary learning: An investigation of L1 and L2 readers' eye movements. *Language Learning*, *71*, 162–203.

Pellicer-Sánchez, A., Siyanova-Chanturia, A., & Parente, F. (2022). The effect of frequency of exposure on the processing and learning of collocations: A comparison of first and second language readers' eye movements. *Applied Psycholinguistics*, *43*(3), 727–756.

Perfetti, C.A. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, *11*, 357–383.

Pulido, D. (2007). The effects of topic familiarity and passage sight vocabulary on L2 lexical inferencing and retention through reading. *Applied Linguistics*, *28*, 66–86.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, *62*, 1457–1506.

Rayner, K., & Duffy, S. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory and Cognition*, *14*(3), 191–201.

Rayner, K., Raney, G.E., & Pollatsek, A. (1995). Eye movements and discourse processing. In R.F. Lorch & E.J. O'Brien (Eds.), *Sources of coherence in reading* (pp. 9–36). Erlbaum.

Reichle, E.D., & Perfetti, C.A. (2003). Morphology in word identification: A word experience model that accounts for morpheme frequency effects. *Scientific Studies of Reading*, *7*, 219–237.

Rice, C., & Tokowicz, N. (2020). A review of laboratory studies of adult second language vocabulary training. *Studies in Second Language Acquisition*, *42*, 439–470.

Robinson, P. (1995). Attention, memory, and the "noticing" hypothesis. *Language Learning*, *45*, 283–331.

Rodd, J. M. (2020). Settling into semantic space: An ambiguity-focused account of word-meaning access. *Perspectives on Psychological Science*.

Rott, S. (1999). The Effect of Exposure Frequency on Intermediate Language Learners' Incidental Vocabulary Acquisition and Retention through Reading. *Studies in Second Language Acquisition*, *21* (4), 589 – 619.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, *11*, 129–158.

Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *The Modern Language Journal*, *95*, 26–43.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, *21*, 360–407.

Torgesen, J.K., Wagner, R., & Rashotte, C. (2012). *TOWRE 2: Test of word reading efficiency*. Pro-Ed Inc.

Wang, A., & Pellicer-Sánchez, A. (2023). Combining eye-tracking and verbal reports in vocabulary research: Benefits and challenges. *Research Methods in Applied Linguistics*, *2*(3), 100063.

Williams, R.S., & Morris, R.K. (2004). Eye movements, word familiarity, and vocabulary acquisition. *European Journal of Cognitive Psychology*, *16*, 312–339.

Cite this article: Elgort, I., & Beyersmann, E. (2024). Do reading times predict word learning? An eye–tracking study with novel words. *Studies in Second Language Acquisition,* 46: 1282–1297. https://doi.org/10.1017/S0272263124000585