




ORIGINAL ARTICLE

The effect of context on the processing and learning of novel L2 vocabulary while reading

Radha Chandy¹ , Raquel Serrano¹  and Ana Pellicer-Sánchez² 

¹Department of Modern Languages and English Studies, University of Barcelona, Barcelona, Spain and

²Department of Culture, Communication & Media, UCL Institute of Education, University College London, London, UK

Corresponding author: Radha Chandy; Email: radhachandy@ub.edu

(Received 23 April 2024; revised 18 September 2024; accepted 4 November 2024; first published online 06 January 2025)

Abstract

Encountering new words multiple times in the input is crucial for incidental vocabulary acquisition. While there is extensive research exploring the impact of word frequency on both learning and processing of novel vocabulary during reading, there is a notable gap in studies examining how contextual factors impact these processes, especially when reading texts, rather than short sentences. The present study aims to fill this gap by exploring the effect of contextual diversity or sameness on adult L2 English learners' processing and incidental learning of novel lexical items through repeated reading of complete texts. Participants ($N = 42$) read one short story three times as well as three different stories, while their eye movements were recorded. Each contextual condition (Same vs. Different) contained ten pseudowords, repeated six times across the treatment. Participants were tested on both immediate and delayed vocabulary learning via form and meaning recognition tests. Our results indicate that repeated readings of the same text led to faster processing as well as better short-term learning of novel vocabulary, although this advantage was not retained for long-term learning. In contrast, initial encoding and lexical integration took longer in the Different condition, although this was not reflected in higher vocabulary gains either in the short- or the long term.

Keywords: Contextual diversity; eye tracking; incidental vocabulary learning; online processing; repeated reading

Reading has been well established as a source of vocabulary learning, in both L1 (e.g., Nagy et al., 1987; Nagy et al., 1985) and L2 research (e.g., Alotaibi et al., 2022; Elgort & Warren 2014; Godfroid et al., 2018; Pellicer-Sánchez & Schmitt, 2010; Webb & Chang, 2012). Although explicit teaching of new vocabulary is often recommended (Nation, 1990, 2001) and intentional learning is usually considered a more effective learning method (e.g., Laufer, 2005; Sonbul & Schmitt, 2010), studies have found that learners of English require knowledge of at least 6,000–7,000 word

© The Author(s), 2025. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

families to understand spoken language and at least 8,000–9,000 in order to read authentic texts (Nation 2001; Schmitt 2008, 2010). It is unlikely that a regular language classroom can provide this volume of knowledge through instruction alone (e.g., de Groot & van Hell, 2005; Horst, 2005) and explicit word-focused activities do not necessarily guarantee learning (Webb et al., 2020), which makes it crucial to study the role of incidental learning in vocabulary acquisition (e.g., Chen & Truscott, 2010; Horst et al., 1998; Pigada & Schmitt, 2006; Pellicer-Sánchez, 2016).

In order for incidental learning to take place, repeated exposure to the target words is required (Nation, 2013; Uchihara et al., 2019). Such repeated exposure to new vocabulary can be promoted by reading the same text multiple times or by reading different texts that include the same target words. Current evidence suggests that encountering new lexical items in a variety of contexts might be particularly beneficial for vocabulary learning (e.g., Adelman et al., 2006; Bolger et al., 2008; Liu & Todd, 2016; Pagán & Nation, 2019; Webb, 2008). Contextual diversity appears to facilitate the encoding and retention of previously unknown words to a greater extent than frequency of exposure (Jones et al., 2017), which suggests that this factor may have a greater influence on vocabulary learning than number of exposures alone. Notwithstanding, the effects of context remain understudied in comparison to investigations on the effects of frequency, a gap which our study aims to address.

In addition, a growing body of research incorporating eye tracking has shown the effect of repeated encounters on the processing of novel vocabulary, generally showing a decrease in reading times with repeated exposures (Godfroid et al., 2018; Elgort et al., 2018; Pellicer-Sánchez 2016; Serrano & Pellicer-Sánchez, 2024). Recent studies have also suggested that reading times seem to be related to vocabulary gains from reading (Godfroid et al., 2013; Koval, 2019; Pellicer-Sánchez, 2016). However, the role of contextual diversity on processing patterns has received little attention. Our study addresses this gap by being the first to explore the role of context in attention to unknown vocabulary in repeated reading of longer texts and its relation to vocabulary learning gains.

Literature review

Incidental vocabulary learning through reading

Although definitions of incidental learning vary, the process may be best understood as a “by-product of meaning-focused activities or tasks” (Webb, 2020, p. 226). This implies that while completing an activity where the focus is to comprehend the content—whether reading, listening, or viewing—lexical knowledge can improve *incidentally* (e.g., Hulstijn, 2001; Schmitt, 2010). However, acquiring both breadth (Nation, 2006a) and depth (Nation, 2001) in L2 lexical knowledge requires extensive exposure to the target forms. Models of lexical access also assume that repeated encounters help fix an item in memory, making it both easier to access and process (Johns, Dye & Jones, 2016).

The key prerequisites for incidental learning to occur via reading are repetition and the possibility of inferring meaning (Uchihara et al., 2019; Webb, 2020). Incidental lexical learning from reading therefore requires repeated and sufficient

exposure to the target items (Pellicer-Sánchez & Schmitt, 2010; Schmitt & Schmitt, 2020), with figures ranging from six (Rott, 1999), to eight (Horst *et al.*, 1998), ten (Pellicer-Sánchez & Schmitt, 2010) or even up to 20 (Waring & Takaki, 2003) exposures.

One reason for such discrepancies may be differences in how frequency (or word repetition) is operationalized in different studies (see Reynolds & Wible, 2014, for a detailed analysis). Another reason could be that different frequencies are needed for learners at differing levels of proficiency (e.g., Horst 2000; Zahar *et al.*, 2001) or for learning and retaining different aspects of lexical knowledge such as receptive versus productive knowledge (Peters, 2020). As a result, while frequency of exposure remains a topic of interest, it is also crucial to focus on how frequency can be moderated by other factors that influence incidental vocabulary learning. Key among these are learner-related variables such as proficiency and text-related variables like the context in which target items appear (Beck *et al.*, 1983; Uchihara *et al.*, 2019; Waring & Takaki, 2003; Webb & Chang, 2015a).

The role of contextual diversity in incidental vocabulary learning

From a theoretical perspective, the positive effects of contextual diversity on vocabulary learning have been explained in different ways. An early explanation is the principle of “likely need” (Anderson & Milson, 1989; Anderson & Schooler, 1991), which relates lexical access to contextual relevance and past access, suggesting that words encountered in a variety of contextual conditions are coded as more likely to be needed in an *unknown* future context, making them more readily accessible in lexical and semantic memory. Bolger *et al.* (2008) explained the benefits of contextual diversity using an “instance-based resonance framework of incremental word learning model” (p. 144; based on Reichle & Perfetti’s 2003 episodic memory model), suggesting that contextual diversity allows words to be encoded in incremental resonant memory traces that include each different context, thereby allowing better consolidation and integration into semantic memory, and the eventual decontextualized encoding of meaning. Another explanation of the benefits of contextual diversity can be found in the semantic distinctiveness model or SDM (Johns & Jones, 2008; Johns, Dye & Jones, 2014; Jones, Johns & Recchia, 2012), which works on the assumption that when a word is encountered in a context that differs from those already encoded in memory, the word is consequently more strongly encoded (Jones *et al.*, 2017).

The role of context has been empirically analyzed from different perspectives, with some studies focusing on the number of distinct contexts in which novel words appear and others focusing on the quality of the contexts in terms of the information they provide: transparency of word meanings, imageability, etc. (see Jones *et al.*, 2017 for a comprehensive summary). The remaining of this section will focus on research into contextual diversity focusing on *number* of contexts, which are the most relevant for the present study.

One of the early studies on the influence of contextual diversity, operationalized as document count, was carried out by Adelman and colleagues in 2006. They compared the effects of word frequency and contextual diversity on the reaction times (RTs) in word naming and lexical decision tasks among English L1 speakers.

They found that contextual diversity influenced item-based RTs to a greater extent than frequency. Results showed that even after accounting for the effects of semantic factors, greater contextual diversity led to faster RTs where frequency did not.

Bolger et al. (2008) explored the role of context by presenting adult L1 speakers of English with rare words either in a single sentence repeated four times or in four different sentences, and either with or without a dictionary definition. They found an effect of contextual diversity on subsequent meaning generation and sentence completion tasks, with the provision of dictionary definitions reducing its effect. The authors conclude that encountering a word in a variety of contexts rather than a single repeated context aided the learning of word meanings, particularly in the absence of dictionary definitions. This would suggest that in situations that require meaning inference (i.e., with no dictionary definitions), contextual diversity has an important impact on vocabulary learning.

In a similar exploration of the effects of contextual variety, this time on L2 learners of English, Ferreira and Ellis (2016) trained participants on real but obscure novel words in either two different sentences or 12 different sentences, while controlling for the frequency of exposures. Participants were subsequently tested on the novel items in a semantic decision task and a read-aloud task. The results indicated that greater contextual diversity led to greater accuracy and faster RTs, as well as faster naming latencies in the post test, indicating that, as reported in Adelman et al., 2006, repeated exposures to novel items in a greater number of contexts facilitate faster processing.

Liu and Todd's 2016 study extended the analysis of contextual diversity to longer texts in the context of repeated-reading practice, exploring the effects of encountering target items in either the same text read seven times ("repeated-reading-same") or in seven different texts ("repeated-reading-different"). Their findings indicate that reading different texts one after the other may lead to more vocabulary gains—as assessed in a meaning recognition multiple-choice test—than reading the same text seven times. The authors suggest that this may be due to the variety of contextual clues provided by different texts benefitting the cognitive processes that lead to learning by "opening up a contextual need" (Zahar et al., 2001, p. 556). They also make a connection between their findings and the observation that learners tend to process input primarily for meaning (VanPatten, 2008), which may lead them to pay less attention to specific words when these words are repeated multiple times in the same context. Additionally, the high number of repetitions in the "repeated-reading-same" condition may have led learners to lose interest with the consequent decrease in attention, although the absence of processing data means that we can only speculate on this point.

Overall, the available studies generally indicate that greater variety of contexts facilitates vocabulary learning both among L1 and L2 speakers, as evidenced in greater accuracy and faster RTs in a number of different tasks (e.g., Adelman et al., 2006; Bolger et al., 2008; Ferreira & Ellis, 2016; Liu & Todd, 2016; Pagán & Nation, 2019), although this may be more beneficial for form recognition than for meaning recognition (Johns et al., 2016). As Jones et al. (2017) suggest, while frequency and contextual diversity may be highly correlated (e.g., Adelman et al., 2006), the number of diverse texts in which a word occurs may be better than just frequency at explaining how new words are learnt and remembered. Furthermore, a greater

range of texts allows for a greater diversity of factors affecting learning such as motivation and interest in the text (Webb, 2020). Notwithstanding these general trends, given the wide variety of participants' L1s, input types, and methods used in previous studies, definitive claims about the benefits of contextual diversity for vocabulary learning cannot be made without further research.

In fact, conflicting findings have also been reported in the literature. When exploring the effects of contextual diversity on incidental vocabulary acquisition through reading, some of the existing studies manipulating semantic diversity have found faster RTs in form recognition tasks for “diverse” contexts, but a benefit of sameness on semantic representations on meaning recognition (e.g., Jones *et al.*, 2012; Johns *et al.*, 2016). As a possible explanation, Johns *et al.*, (2016) suggest that while semantic diversity aids the stronger encoding of novel forms, it may hinder comprehension and meaning inference, which is easier in semantically constant contexts. Similarly, other studies have also shown that contextual sameness may be more beneficial than diversity, especially for learners with lower comprehension skills or proficiency (e.g., Elgort *et al.*, 2015; Elgort & Warren, 2014; Horvath & Arunachalam, 2021; Zahar *et al.*, 2001). However, there are not enough studies to draw definitive conclusions on how context affects incidental vocabulary acquisition—especially among L2 learners—and thus the aim of the present study.

Online processing and vocabulary learning in SLA

The examination of eye movements is an important method to better understand the cognitive processes involved in reading and learning vocabulary under different conditions (Pellicer-Sánchez, 2020), while coming as close as experimentally possible to capturing the natural processes of silent reading (Pagán & Nation, 2019). Studies on incidental learning emphasize the need to understand the amount and kind of attention novel items receive in order to better explain factors affecting vocabulary gains (e.g., Nassaji, 2006; Pulido, 2007). The time spent processing a new linguistic element can be considered a measure of the amount of cognitive engagement with the material (Godfroid *et al.*, 2013, p. 486).

Previous eye-tracking studies have shown that readers spend more time looking at (i.e., show longer fixations on) novel words than words they already know (e.g., Godfroid *et al.*, 2018; Pellicer-Sánchez, 2016). Furthermore, within L2 reading research, eye-tracking has shown that reading times decrease across repetitions (e.g., Elgort *et al.*, 2018; Mohamed, 2018; Pellicer-Sánchez, 2016, Pellicer-Sánchez *et al.*, 2022), suggesting increasingly fluent processing of the text as a function of repeated encounters. In contrast, longer processing times on target vocabulary have been linked to better learning outcomes in some studies (Godfroid *et al.*, 2013; Koval, 2019; Mohamed, 2018; Pellicer-Sánchez, 2016) but not in others (e.g., Wochna & Juhasz, 2013), calling for more research in this area.

Crucially, the patterns reported by the studies above are bound to be influenced by the context in which novel words appeared. Considering that learners tend to process textual input primarily for meaning (VanPatten, 2008), encountering novel items in the same context repeatedly could lead to a decrease in reading times, since understanding a word may not be necessary to furthering global comprehension of the text (Liu & Todd, 2016). Encountering the novel words in diverse contexts,

however, would require learners to process them for meaning each time, potentially leading to increased attention and greater gains. However, to our knowledge, only one study has investigated the effect of contextual diversity on attention to novel words in reading.

Pagán and Nation (2019) measured the eye movements of L1 speakers of English while encountering novel words in one repeated sentence (same context) against four different sentences (diverse context). Accurate responses to comprehension questions testing understanding of the target items were taken as evidence of vocabulary learning, as were faster processing times (measured by eye-movements to target items) in the post-task phase as compared to the experimental phase. Their findings showed that words read in the same context received shorter viewing times than words read in the diverse context—an effect that was visible in both early (first fixation and gaze duration) and late (go past and total time) processing measures. In the post-task testing phase, however, words that had appeared in the diverse condition received shorter fixations, indicating better learning and integration in the lexicon. The authors explain these patterns in terms of contextual diversity allowing representations of the words in the learners' minds to become more context independent (Pagán & Nation, 2019). Crucially, similar investigations in the L2 context are yet to be conducted. Our study investigates whether similar processing patterns and learning outcomes may be observed when learners encounter target items embedded in longer texts, comparing the same text read multiple times with the reading of different texts. Moreover, apart from investigating processing during reading, our study also examines whether differences in processing are connected to differential lexical gains, as observed in vocabulary tests.

This study

As the review of the literature shows, besides the general scarcity of research on contextual diversity in the L2 context, research exploring the effects of contextual diversity on vocabulary learning (e.g., Bolger et al., 2008; Ferreira & Ellis, 2016) and studies exploring its effect on processing of new vocabulary (Pagán & Nation, 2019) have tended to focus on novel items embedded in single sentences. However, learners are often required to read longer texts for meaning, and it is important to understand what factors affect the learning as well as processing of novel words in longer texts. Therefore, in view of preliminary findings suggesting the benefits of contextual diversity for vocabulary learning in such conditions (e.g., Liu & Todd, 2016), our aim is to investigate both processing and incidental vocabulary learning through the repeated reading of longer texts, comparing contextually same versus different conditions.

While several reading studies have used eye tracking to explore incidental vocabulary acquisition while reading longer texts (e.g., Godfroid et al., 2013; Mohamed, 2018; Pellicer-Sánchez et al., 2022), none, to our knowledge, have combined the study of processing and incidental learning through repeated reading of longer texts in the L2 while comparing same vs. different contexts. Understanding how externally manipulated conditions such as contextual diversity can affect how novel items are processed and to what extent they are learnt also has important

pedagogical implications for the design of classroom reading tasks, for example. Finally, insights from processing may help explain the potential effects of contextual diversity on incidental vocabulary learning.

The present study therefore aims to investigate how contextual diversity (in the form of multiple texts) vs. contextual sameness (the same text read multiple times) affect the learning and processing of novel lexical items in L2 repeated reading and to what extent processing patterns may explain how well these items are learnt when the number of exposures (i.e., frequency of occurrence) is kept constant. To that end, our research questions are as follows:

1. Are there differences in the processing of novel vocabulary under same versus different repeated reading conditions?
2. Do vocabulary learning and retention differ when new words are repeated in the same text read multiple times versus different texts?
3. Do processing patterns explain vocabulary learning and retention under different contextual conditions?

Based on the existing literature, we can formulate certain hypotheses. In response to the first research question, we expect reading times to decrease across repetitions (e.g., Elgort *et al.*, 2018; Godfroid *et al.*, 2018; Mohamed, 2018; Pellicer-Sánchez 2016). We might also expect greater decreases in fixation times in the Same condition as opposed to the Different condition in line with episodic memory models (Bolger *et al.* 2008; Reichle & Perfetti's 2003), the SDM and with Pagán and Nation's (2019) findings, since repetitions in the Different condition would require increased processing effort in order to be encoded in memory. Greater decreases in fixations in the Same condition would also be predicted by Liu and Todd's (2016) theory of "input processing predilection" (p. 69).

In terms of vocabulary gains (RQ2), we might hypothesize greater vocabulary gains under the Different condition (Bolger *et al.*, 2008, Ferreira & Ellis, 2016; Liu & Todd, 2016; Pagán & Nation, 2019). Concerning the third research question, as indicated in the literature review, we can speculate that longer processing times might be indicative of greater cognitive processing of the items, which, in turn, might lead to more learning (e.g., Godfroid *et al.*, 2013; Mohamed, 2018; Pellicer-Sánchez, 2016).

Methods

Participants

The participants were 42 undergraduate students enrolled in the English Studies undergraduate degree at a university in Catalonia, Spain (five males and 37 females, aged between 19 and 23 years old). Twenty-seven were L1 Spanish/Catalan bilingual, 11 were multilingual (heritage speakers of another language along with Spanish/Catalan), and four were Erasmus exchange students from other parts of Europe (L1 Italian $n = 2$; German $n = 1$; Polish $n = 1$). All participants were enrolled in an Applied Linguistics class, and they received course credit for voluntarily taking part in the study. All participants had a vocabulary size above

Table 1. Descriptive statistics for vocabulary size and proficiency by L1 group

L1	N	Vocabulary Size		Quick OPT Score	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Catalan, Spanish	27	6666.29	1015.14	46.26	5.84
Catalan, Spanish, other/another language(s)	11	6628.64	1477.3	47.27	6.23
Other European Languages (Exchange Students)	4	6723.25	1057.1	47	6.75

Notes: Max. score for vocabulary size = 10000. Max score for OPT = 60

4,000 according to V_YesNo Test (Meara & Miralpeix, 2017): $M = 6661.86$, $SD = 1157.82$. In addition, according to the Quick Oxford Placement Test (UCLES, 2001), the participants had an average score of 46.6/60 ($SD = 6.05$), indicating a B2 or upper intermediate proficiency level. As Table 1 illustrates, the scores of the vocabulary and grammar tests across L1 groups were very similar.

Reading materials

Four stories were designed for the purpose of this study. One story (henceforth Text 1) was read three times in the Same condition, while the other three stories (henceforth Texts 2–4) were read one after the other in the Different condition. Following the operationalization of contextual diversity as document count (e.g., Ferreira & Ellis, 2016; Jones et al., 2017; Liu & Todd, 2016; Pagán & Nation, 2019), in the present study, we considered target items to be in the same context when they appeared in the text that was read three times (Text 1) and in a different context when they appeared in Texts 2, 3, and 4, each of which were only read once. Each story consisted of 16 paragraphs and contained 10 pseudowords, with each word repeated twice within the text, resulting in a total of six repetitions, in line with the minimum recommended by Rott (1999). We decided to include two repetitions in each text and not just one occurrence in order to limit the number of re-readings of the entire text to three, considering the possibility that six repeated readings might lead to reduced attention due to boredom.

The text assigned to the Same condition (Text 1) was adapted from a text used in a previous study by Pellicer-Sánchez (2016). This text was shortened to include only two repetitions of the target items instead of the original eight, given that our participants read the text multiple times as opposed to one. The three texts used in the Different condition (Texts 2–4) were simple fictional narratives told in the past tense in a similar style to Text 1, taken from the website *storynory.com* and adapted to contain the same number of repetitions of the target items in each.

To ensure that the texts would be easily comprehensible to L2 participants and to facilitate noticing and processing of new lexical items, the words used in all four stories were high frequency words, mostly belonging to the first 2,000 most frequent English word families. Knowledge of the first 3,000 word families provided 98% lexical coverage. Given the average level of proficiency of participants (i.e., B2 or upper intermediate) and their overall vocabulary size ($M = 6,572$), we expected them to have no major issues comprehending the texts. The four texts all had a

Table 2. Text characteristics

	Text 1	Text 2	Text 3	Text 4
Number of words	953	955	953	947
Words/sentence	11.2	10.6	11.5	12.3
Word frequency	K1-K3	K1-K3	K1-K3*	K1-K2
Coverage 95%	K2	K2	K2	K2
Coverage 98%	K3	K3	K3	K3
Flesch-Kincaid Grade Level	4.1	3.9	4.2	4.5
Flesch Reading Ease score	85.5	86.1	85.2	84.4
Reading Level	6th grade	6th grade	6th grade	6th grade

*There was one K5 word, *jealous*, which is a cognate with Catalan/Spanish *gelós/celoso* and was assumed to be known by the target participant.

similar lexical profile and readability scores (see Table 2 for text characteristics and Appendix A in the supplementary materials for the complete texts).

The stories were presented on the desktop-based Tobii Pro Spectrum eye-tracking monitor (23.8" diagonally, 1920 × 1080 pixel resolution) as black text on a white screen in Courier New font, size 18, with double line spacing. Each text was divided into sixteen screens, with one paragraph of the story and a maximum of 69 words per screen ($M = 63.53$). No more than two (different) target words were present on any one screen, and the target items were not placed at the beginning or end of a line or sentence since existing research indicates that the first fixation on a line may be longer (Rayner, 1977). All participants read all four texts, with half of them randomly assigned to either a Same-first or Different-first group, whereby approximately half read the text in the Same condition first, while the other half read the texts in the Different condition first.

Target items

The target items for the study were 20 nonwords chosen from the ARC Nonword Database (Rastle, Harrington, & Coltheart, 2002) and used to replace 20 countable, high-frequency nouns from the texts (within the first 1K or 2K word families). They were selected based on the following criteria: neighborhood size (min = 1, max = 5), number of body neighbors (min = 1, max = 5), and number of phonological neighbors (min = 1, max = 5). In addition, no pseudo-homophones were used, and only orthographically existing onsets, orthographically existing bodies, and legal bigrams were extracted.

A pool of pseudowords was first piloted with L1 speakers of English, who were asked to rate 63 pseudowords intermixed with 12 low-frequency real words (e.g., *burlap*, *scab*, and *churn*) on a six-point Likert scale according to how English-like, or real word-like, their form was (1 = very unlikely to be an English word, and 6 = very likely to be an English word). Based on the scores provided, we selected 20 pseudowords whose mean rating was at least 4. We ensured that all the items that

replaced plural nouns also finished with a regular plural morpheme (-s, -es), and that none of the items that replaced singular nouns ended in -s or -es. All were single syllable words (e.g., *glabe*, *brald*) and their lengths varied between 4 and 7 letters. For details of the pseudoword characteristics see Appendix B, and for pseudoword target meanings and distribution within the texts are see Appendix C (both included in the supplementary materials).

Even though we made sure the target pseudowords were similar with respect to features that might affect their processing and learning, as reported above, we also took a further step to control for the role of the target pseudowords. The 20 nonwords were divided into two sets of 10 words each, with approximately half of the participants being exposed to set 1 in Text 1 (Same condition) and set 2 in Texts 2–4 (Different condition) and the other half reading set 2 in Text 1 and set 1 in Texts 2–4 (see Appendix D in the supplementary materials for details of the experimental procedure, including the counterbalanced format). To control for the degree of informativeness of each specific context, 31 L1 speakers of English were asked to assess the likelihood of inferring the meaning of the target pseudowords by considering the two sentences in which they appeared using a 4-point rating scale similar to Webb's (2008):

- 1 = Extremely unlikely that the target word can be guessed correctly.
- 2 = It is unlikely that the meaning of the target word can be inferred.
- 3 = It might be possible to infer the meaning of the target word.
- 4 = Readers have a good chance of guessing the meaning of the target word.

The mean score for all 20 pseudowords across the 31 raters was 3.31 ($SD = .246$), indicating that meanings were relatively easy to infer from all the contexts in which the items appeared. Crucially, no significant differences were found in contextual informativeness between words that appeared in the Same (Text 1) and Different (Texts 2–4) conditions ($t(38) = .636$, $p = .528$).

Vocabulary tests

To test different aspects of target item knowledge, we used three different vocabulary tests. First, in a form recognition test, participants were presented with a list of the 10 target items that appeared in each condition mixed with 10 distractor pseudowords and were told to choose any words they remembered seeing in the readings (no minimum or maximum limit was placed). The distractor pseudowords were also taken from the ARC Nonword Database (Rastle, Harrington, & Coltheart, 2002), following the same criteria as the target pseudowords (see Appendix E in the supplementary materials for a list of the distractors). The form recognition was followed by a meaning recall test, in which participants were presented with the set of pseudowords and asked to provide a meaning (in the form of a synonym, short definition, or translation in Spanish/Catalan). Participants finally completed a multiple-choice meaning recognition test, which was expected to capture more learning gains than the meaning recall test (Waring & Takaki, 2003). The test contained three options: the target meaning, two distractors either taken from within the stories or that were plausibly semantically related to the content of the stories, and the option “I don't know”. The three tests were completed in the same

order via Google Forms (see Appendix F in the supplementary materials for the complete vocabulary tests).

Procedure

All participants were exposed to both contextual conditions and all 20 pseudowords in a counter-balanced, within-subjects design. The procedure consisted of three sessions, each lasting approximately half an hour to one hour. Prior to beginning the study, all participants were informed about the general aims of the study and signed an informed consent form. Additionally, participants filled in a language-background questionnaire which included questions on their L1(s) and experience learning and reading in English. In the first session, participants filled out the language background questionnaire, and then read three texts consecutively: either Text 1 three times or Texts 2–4, depending on whether they were in the Same-first ($n = 23$) or Different-first ($n = 19$) group. Since we were interested in incidental vocabulary learning and in order to ensure that participants were focusing on meaning, the texts were followed by 16 true-false questions about the main ideas in each paragraph of the text(s) (see Appendix G in the supplementary materials for complete comprehension tests). The participants were informed about these comprehension questions before they started reading, and they answered them on a different computer via Google Forms. None of the target items appeared in any of the questions.

Participants read the texts on a desktop-mounted Tobii Pro Spectrum monitor, which records eye movements at a maximum sample rate of 1200 Hz through the Tobii Pro Lab software (Tobii Pro AB, 2014). Participants were seated 55–75 cm from the monitor, as recommended for optimal recording of eye movements. Before each session, accuracy was verified using a five-point calibration and validation. If calibration values were $> 0.5^\circ$, the procedure was repeated. Participants read the texts silently and were told to minimize head movement. They navigated from one screen to the next by mouse-click and could therefore control their own pace. They were not allowed to go back to previously read screens. Participants took around 10–12 minutes to complete the readings in the Same condition and 15–18 in the Different condition.

Session 2 was held after a one-day gap. In this session, participants who had started with the Same condition read texts in the Different condition and vice versa. Once again, they were told to read for meaning, and the readings were followed by true/false comprehension questions. Overall, comprehension scores showed an average of 13.87/16 in the Same condition (86.69% correct responses) and 12.04/16 in the Different condition (75.25% correct responses) suggesting that participants did not have major comprehension difficulties. This session also included the vocabulary posttests of the words encountered in Session 2 readings, which were used to obtain a measure of participants' immediate gains. In order to control for potential test effects, the immediate test was only conducted after the second session and included the items presented in that session. Before completing the immediate vocabulary posttests, participants did the V_YesNo Vocabulary size test (Meara & Miralpeix, 2017), which took approximately 10 minutes.

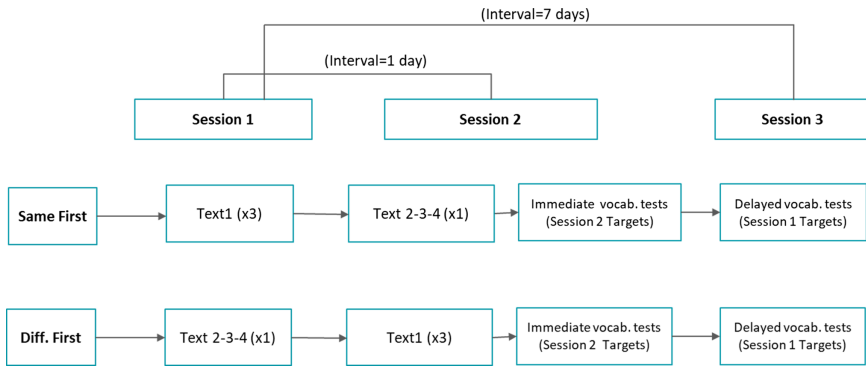


Figure 1. Schematic layout of the experimental design.

The final session, held a week after the first, consisted of delayed vocabulary posttests corresponding to the items encountered in the Session 1 readings. The tests were again completed in fixed order (1. form recognition, 2. meaning recall, and 3. meaning recognition). This was followed by the Quick OPT (UCLES, 2001), which was used to obtain a measure of participants’ English proficiency. As described in this section, and in line with some previous studies (e.g., Rogers & Cheung, 2021), participants were only tested on each pseudoword once to avoid test repetition effects. Moreover, administering vocabulary tests after the first reading session would probably have increased participants’ focus on the new words during Session 2, affecting the incidental learning nature of the study. Figure 1 presents a summary of the experimental design.

Analyses

Processing measures

Areas of interest were created around each target item, with a uniform height of 100 pixels and a length varying between 105 and 150 pixels, depending on the length of the pseudoword. We then extracted fixation measures for each separate encounter with each pseudoword (six encounters over three readings for all pseudowords and conditions). Tobii’s I-VT fixation filter was applied to extract the data, which resulted in fixations lower than 60 ms being discarded.

The eye-tracking measures extracted were the following: total fixation duration (TFD), which is the sum of the durations of all fixations on an item; average fixation duration (AFD), which measures the average length of each fixation on the target area; first fixation duration (FFD), which is the initial or first fixation on the item; and fixation count (FC), which refers to the total number of fixations on the target item. FFD is considered an “early” measure that captures recognition and lexical access (Godfroid, 2019), while TFD and FC are “later” measures that reflect lexical integration (Rayner, 2009), as well as the contextual- and discourse-level properties of what is being read (Conklin et al., 2018). AFD, meanwhile, may indicate how eye movements evolve over time (Conklin et al., 2018).

Vocabulary scores

The tests for vocabulary form and meaning recognition were scored dichotomously. On the form recognition test, one point was awarded for each target word accurately recognized; on the meaning recognition test, students received one point for each target pseudoword that was correctly matched to the original word from the texts it replaced. For meaning recall, answers were scored as 1 when students were able to either 1) provide the original words the target pseudo-words replaced or a synonym or word that was accurate in the target context (e.g., *plates* instead of *bowls*, or *path* instead of *road*), 2) provide a translation in Spanish/Catalan, and 3) or define the meaning of the target pseudowords. Answers that did not meet either criterion were scored as 0. Two raters scored the answers and reached 100% agreement on the scores.

Statistical analyses

Version 27.0 of the SPSS program (IBM Corp, 2020) was used for the statistical analyses. Since data from the eye-tracking measures were not normally distributed and were skewed toward higher values, *a-priori* cleaning was carried out by eliminating all values that were 2.5 SD above the mean (as recommended in Godfroid, 2019). The resulting data loss were as follows: TFD 3.02%, AFD 1.47%, FFD 1.96%, and FC 2.58%. The remaining data were normally distributed according to the skewness and kurtosis values, which were between -2 and 2 (Godfroid, 2019) for all four measures. In terms of the distribution of vocabulary scores, results for the form and meaning recognition tests also presented skewness and kurtosis values between -2 and 2 again indicating normal distribution (see Appendix H in the supplementary materials for normality all skewness and kurtosis values).

To examine processing (RQ1), we performed repeated measures linear mixed models (LMM) with participants as random factors, item number, and encounter time (1 and 2 for the two repetitions of each item in teach text) as repeated measures and the eye-tracking measures (TFD, AFD, FFD, and FC) as dependent measures. Initially, participants' L1 was included as a fixed effect in the model but since it did not have a significant effect,¹ the factor was removed and the model was run again with only condition (Same vs. Different), reading time (1, 2, 3), and their interaction introduced as fixed effects. The SPSS model adjusted *p*-values for multiple comparisons using sequential Bonferroni. Standardized Pearson residuals were saved and were within the normal range for all the target factors (no absolute values greater than 2, as recommended in Godfroid, 2019) reducing the likelihood of outliers affecting model fit.

In order to answer RQ2, generalized linear mixed models (GLMM) with repeated measures and participants as random factors were run, including the scores in the form and meaning recognition vocabulary tests as dependent measures. Condition (Same vs. Different) and time (immediate vs. delayed posttest) and their interaction were included as fixed effects. Finally, to examine the effects of processing on vocabulary learning (RQ3), the same model was re-run, this time with TFD added as a fixed effect. Only one eye-tracking measure was selected due to the strong correlation between them. TFD was chosen as it is the measure most strongly

associated with learning in existing research (Godfroid et al., 2018) and the one that best captures global processing patterns (Godfroid, 2019). Estimated marginal means were obtained for each categorical factor, with p -values of pairwise comparisons corrected using the sequential Bonferroni method. Fixed coefficients were obtained for continuous variables. Once again, standardized Pearson residuals were saved and found to be within the normal range for all of the target factors.

SPSS was unable to perform the GLMM analysis with the scores for meaning recall, due to the very high percentage of zeros in the results. From a total of 840 cases (20 target words \times 42 participants), 738 (87.9%) were incorrect (0) and 102 (12.1%) were correct (1).

Results

RQ1. Processing patterns

The descriptive statistics of fixations on target items are presented in Table 3. These show a decrease in processing times and number of fixations from the first to the third reading, in both the Same and Different conditions, with a clearer decrease in the former.

The results of the LMMs exploring the fixed effects of condition, reading times, and their interaction are presented in Table 4. The parameter estimates (SPSS fixed coefficients table) are included in Appendix I.

For TFD, there was a significant effect of condition ($p = .030$) and reading time ($p < .001$). Pairwise contrasts between Same and Different condition indicated that fixations were longer in the Different condition ($t(4882) = -2.173, p = .030$). When comparing processing patterns across the three reading times, pairwise contrasts showed that fixations were significantly shorter at Time 2 than Time 1 ($p < .001$), at Time 3 than at Time 2 ($p = .037$) and at Time 3 than at Time 1 ($p < .001$). There was a significant interaction between condition and reading time ($p = .038$). The pairwise comparisons suggest that, while changes in eye-movements across reading times were equivalent for the two conditions, showing a significant decrease (except for Times 2–3 [$p = .140$]), when comparing conditions across reading times, TFD was significantly higher in the Different condition at Time 2 ($p = .016$) and Time 3 ($p = .012$) but not at Time 1 ($p = .395$).

In terms of AFD, there was also a significant effect of condition ($p < .001$) and reading time ($p < .001$). Pairwise contrasts showed a similar pattern to TFD, with average fixations in the Different condition being longer than those in the Same condition ($t(4960) = -4.016, p < .001$). Pairwise contrasts across reading times showed a significant decrease in AFD from Times 1–2 ($p = .001$), Times 2–3 ($p < .001$), and Times 1–3 ($p < .001$). We again found a significant interaction between condition and reading time ($p = .005$). Subsequent pairwise contrasts revealed a significant decrease across all reading times for both conditions, except for Times 2–3 in the Same condition ($p = .256$). When comparing conditions across time, no significant differences were found at Time 1 ($p = .822$) and Time 3 ($p = .051$), although fixations were shorter in the Same condition at Time 2 ($p < .001$).

The results for FFD showed a significant main effect of condition ($p < .001$) and reading time ($p < .001$). Pairwise contrasts between Same and Different conditions

Table 3. Descriptive statistics for fixation measures on target items across the three repeated readings by condition. Fixation durations in milliseconds

ET Measure	SAME								
	Reading 1			Reading 2			Reading 3		
	<i>M</i>	<i>SD</i>	95% CI	<i>M</i>	<i>SD</i>	95% CI	<i>M</i>	<i>SD</i>	95% CI
TFD	512.29	330.97	[488.12, 535.95]	388.97	313.15	[366.62, 411.37]	359.45	309.34	[336.62, 379.47]
AFD	208.12	91.95	[202.06, 214.36]	179.49	98.53	[172.75, 186.23]	175.53	105.09	[167.99, 183.04]
FFD	201.87	98.64	[194.85, 208.99]	180.26	108.50	[172.83, 188.34]	175.61	113.64	[167.58, 183.34]
FC	2.32	1.42	[2.22, 2.42]	1.83	1.39	[1.74, 1.93]	1.68	1.37	[1.58, 1.77]
ET Measure	DIFFERENT								
	Reading 1			Reading 2			Reading 3		
	<i>M</i>	<i>SD</i>	95% CI	<i>M</i>	<i>SD</i>	95% CI	<i>M</i>	<i>SD</i>	95% CI
TFD	493.74	313.76	[470.70, 517.12]	414.60	287.93	[395.48, 434.32]	381.81	271.45	[362.45, 399.43]
AFD	207.35	86.36	[200.99, 213.66]	200.95	87.71	[195.09, 207.14]	186.01	95.66	[179.06, 192.45]
FFD	206.68	100.81	[199.61, 213.87]	198.29	100.04	[191.73, 205.65]	183.09	103.21	[175.80, 190.58]
FC	2.28	1.43	[2.19, 2.39]	1.95	1.30	[1.86, 2.04]	1.85	1.29	[1.76, 1.94]

Notes: TFD = total fixation duration; AFD = average fixation duration; FFD = first fixation duration; FC = fixation count.

Table 4. LMMs. Fixed effects of condition, reading time, and condition*reading time on eye-movement measures

ET Measure	Source	<i>F</i>	<i>df1</i>	<i>df2</i>	Sig.
TFD	Corrected Model	38.485	5	4882	<.001
	Condition	4.722	1	4882	0.030
	Reading Time	89.732	2	4882	<.001
	Condition*Reading Time	3.265	2	4882	0.038
AFD	Corrected Model	27.032	5	4960	<.001
	Condition	16.131	1	4960	<.001
	Reading Time	54.413	2	4960	<.001
	Condition*Reading Time	5.351	2	4960	0.005
FFD	Corrected Model	13.420	5	4935	<.001
	Condition	12.718	1	4935	<.001
	Reading Time	25.499	2	4935	<.001
	Condition*Reading Time	1.629	2	4935	0.196
FC	Corrected Model	32.990	5	4904	<.001
	Condition	5.734	1	4904	0.017
	Reading Time	74.503	2	4904	<.001
	Condition*Reading Time	4.263	2	4904	0.014

again showed that initial fixations were longer in the Different condition than in the Same ($t(4935) = -3.566, p < .001$). In terms of processing across reading times, there was again a significant drop in FFDs from Times 1–2 ($p < .001$), Times 2–3 ($p = .006$), and Times 1–3 ($p < .001$). The interaction between the condition and reading time was not found to be significant ($p = .196$).

Finally, for FC, there was a significant effect of condition ($p = .014$), reading time ($p < .001$) and of the interaction between condition and reading time ($p = .014$). Pairwise contrasts, in line with the results for the other eye-tracking measures, indicated that the number of fixations was higher in the Different condition than in the Same condition ($t(4904) = -2.394, p = .017$). For reading times, pairwise contrasts indicated a significant drop in the number of fixations from Times 1–2 ($p < .001$), Times 2–3 ($p = .038$), and Times 1–3 ($p < .001$). There was a significant interaction between condition and reading time ($p = .005$). The pairwise comparisons showed that while changes across reading times were equivalent in both conditions (significant decrease between Times 1–2 and Times 1–3 but not Times 2–3), there were significantly fewer fixations in the Same condition at Time 2 ($p = .011$) and Time 3 ($p = .003$), although not at Time 1 ($p = .314$).

Details of all significant pairwise comparisons for all four processing measures can be found in Appendix I in the supplementary materials. Figure 2 presents a graphical representation of the interaction between reading time and condition for each of the processing measures.

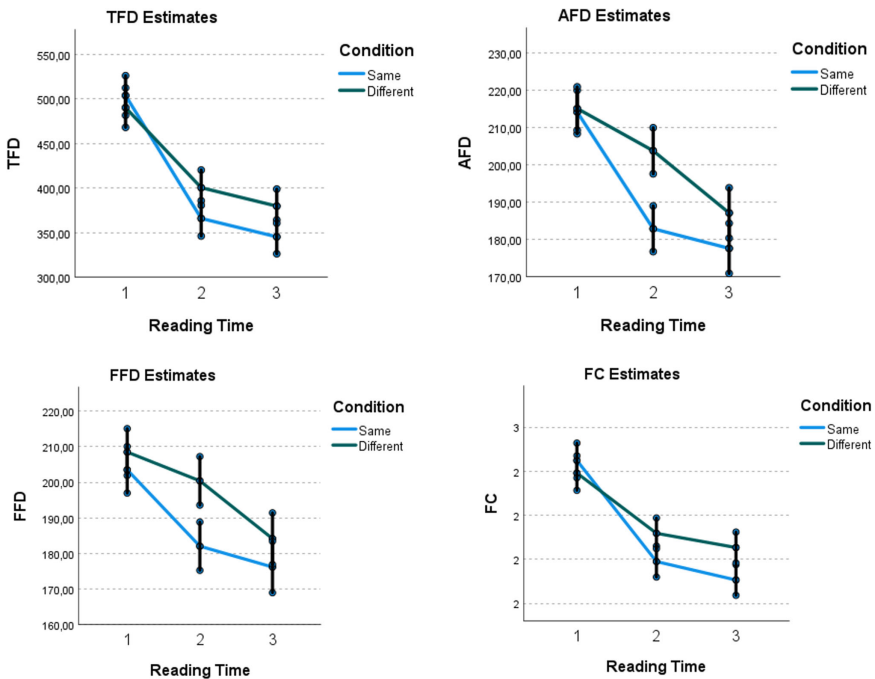


Figure 2. Estimated marginal means for condition*reading time (see Appendix I for complete estimated marginal means tables).

RQ 2. Context and vocabulary scores

The descriptive statistics for vocabulary scores under the Same and Different conditions and by test time (immediate and delayed) are presented in Table 5. In the Same condition, higher mean scores were obtained at the immediate rather than delayed posttest for both form and meaning recognition. In the Different condition, form recognition scores were higher in the immediate posttest, although, interestingly, mean scores were slightly higher in the delayed posttest in the case of meaning recognition. Overall, the scores in the Same condition were higher than in the different condition in the immediate posttest, but this advantage disappears in the delayed posttest.

Results for the GLMMs examining the fixed effects of condition, test time and the interaction between condition and test time are presented in Table 6.

For form recognition, there was a significant effect of test time ($p < .001$), but neither condition ($p = .313$) nor the interaction between test time and condition ($p = .512$) showed significant effects. Pairwise contrasts demonstrated that scores were significantly higher on the immediate posttest than on the delayed posttest ($t(836) = 6.997, p < .001$).

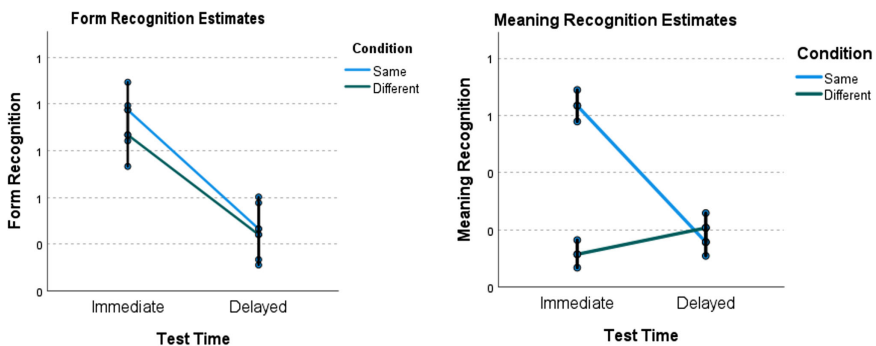
The results for meaning recognition showed a significant effect of both condition ($p = .001$) and test time ($p = .003$). Pairwise contrasts between the Same and Different conditions showed that participants who read the same text three times scored higher than those who read three different texts ($t(836) = 3.446, p = .001$).

Table 5. Descriptive statistics for vocabulary scores in the immediate and delayed posttests by condition. Scores out of 1

Vocabulary Test	SAME						DIFFERENT					
	Immediate			Delayed			Immediate			Delayed		
	<i>M</i>	<i>SD</i>	95% CI	<i>M</i>	<i>SD</i>	95% CI	<i>M</i>	<i>SD</i>	95% CI	<i>M</i>	<i>SD</i>	95% CI
Form Recognition	0.66	0.47	[0.64, 0.69]	0.45	0.49	[0.42, 0.47]	0.61	0.48	[0.58, 0.64]	0.44	0.49	[0.41, 0.47]
Meaning Recognition	0.53	0.49	[0.50, 0.56]	0.31	0.46	[0.28, 0.33]	0.29	0.45	[0.27, 0.31]	0.33	0.47	[0.30, 0.36]

Table 6. GLMMs. Fixed effects of condition, test time, and condition*test time on vocabulary test scores

Test	Source	<i>F</i>	<i>df1</i>	<i>df2</i>	Sig.
Form Recognition	Corrected Model	15.444	3	836	< 0.001
	Condition	1.017	1	836	0.313
	Test Time	45.333	1	836	< 0.001
	Condition*Test Time	0.43	1	836	0.512
Meaning Recognition	Corrected Model	16.188	3	836	< 0.001
	Condition	11.703	1	836	0.001
	Test Time	8.706	1	836	0.003
	Condition*Test Time	23.382	1	836	< 0.001

**Figure 3.** GLMMs: Estimated marginal means for condition*test time.

They also scored higher on the immediate than the delayed posttests ($p = .003$). There was a significant interaction between condition and test time ($p < .001$). Pairwise comparisons examining condition at each test time revealed significantly higher scores in the Same condition at the immediate posttest ($p < .001$), but no significant differences between conditions at the delayed posttest ($p = .171$). Further pairwise contrasts showed that in the Same condition, participants scored significantly higher on the immediate post-test ($p < .001$), although there was no significant difference between the two test times in the Different Condition ($p = .191$). Figure 3 shows a graphical representation of the interactions. See Appendix J in the supplementary materials for details of pairwise comparisons.

RQ3. Relationship between vocabulary and processing

In order to explore how processing relates to vocabulary learning, the GLMM model used for RQ2 was re-run to include processing (measured as TFD) as a fixed effect. We were especially interested in the interaction between TFD and condition (Same vs. Different) and also in the three-way interaction with testing time to examine

Table 7. GLMMs. Fixed effects of TFD and TFD*test time on vocabulary test scores. Only significant results shown

Test	Source	<i>F</i>	<i>df1</i>	<i>df2</i>	Sig.
Form Recognition	Corrected Model	25.808	5	4882	< 0.001
	TFD	5.630	1	4884	0.018
Meaning Recognition	Corrected Model	68.993	3	4884	< 0.001
	TFD	4.376	1	4882	0.036
	TFD*Test Time	4.677	1	4882	0.031

differential effects between conditions for short- and long-term vocabulary learning. Initially, all fixed effects as well as possible two- and three-way interactions were included in the model. The results showed no interactions between TFD and condition (Same vs. Different), suggesting that the relationship between eye-movements and vocabulary learning were not different between Same and Different contexts (see Appendix K in the supplementary materials for all the results).

Another model was run in which these interactions were removed, for a better understanding of the results on how processing relates to vocabulary learning in repeated reading. Table 7 shows significant effects of TFD on form and meaning recognition scores.

In terms of form recognition scores, TFD had a significant overall positive effect (OR = 1.24, 95% CI [1.04, 1.49]). This implies that, across conditions and test times, for each second more in TFD, the odds of correctly recognising the form of the target item went up by 24%.²

For meaning recognition scores, results showed a significant overall effect of TFD (OR = 1.48, 95% CI [1.14, 1.92]) as well as a significant interaction between TFD and test time ($p = .031$). At immediate testing, the effect of TFD (OR = 1.48, 95% CI [1.04, 2.12]), was shown to be higher than at delayed testing (OR = 1.001, 95% CI [0.78, 1.29]). This indicates that one second longer in TFD resulted in 48% greater odds of correct meaning recognition at the immediate posttest, but only 0.1% greater odds at the delayed posttest (see Appendix K in the supplementary materials for complete parameter estimates).

Discussion

This study explored the effect of contextual diversity on the processing and learning of novel vocabulary through repeated reading. A group of undergraduate EFL learners encountered 20 pseudowords (10 per condition), each occurring six times across three readings of either the same or different English texts. Our results show that despite the difficulty of the task—completely unknown words, and only moderate frequency of exposure—participants managed to recognize the form and meaning of several of the target words with some even being able to produce an accurate translation or definition. This, by itself, is testament to the effectiveness of repeated reading for developing new lexical knowledge in the L2 (e.g., Liu & Todd, 2016; Webb & Chang, 2012).

When it comes to processing and reading patterns in the context of a repeated reading design that compares the same versus different texts, our findings show that the time spent processing novel items in repeated readings of the same text was significantly lower than when the texts were different. This is similar to Pagan and Nation's (2019) study, where target items read in the Same-condition sentences showed shorter fixations in both early and late measures of processing. Our data indicates that condition (Same vs. Different) had a significant fixed effect on processing patterns across all four processing measures, although certain differences were observed in each measure studied.

The interaction between condition and reading times was significant for three of the four processing measures (TFD, AFD, and FC). Despite a lack of statistically significant differences in eye movements between conditions during the first reading, the number and duration of fixations in subsequent readings (2 and 3) were significantly lower in the Same condition than in the Different condition. These findings suggest that the integration of contextual and discourse-level properties (Conklin *et al.*, 2018; Rayner, 2009) captured by a reduction in the number (FC) and length (TFD) of fixations across readings is slower when participants re-encounter novel items in contextually diverse texts. The significantly lower average processing time (AFD) in readings 2 and 3, along with the larger mean decreases in AFD from one reading to the next in the Same condition, indicate that overall processing evolves more slowly in the contextually Different condition, given that AFD is considered an evolutive measure (Conklin *et al.*, 2018). These results are also coherent with episodic memory models that suggest increasing contextual diversity would require more processing at each new encounter (Bolger *et al.* 2008; Reichle & Perfetti 2003). They also coincide with the theory of input processing predilection (Liu & Todd, 2016; VanPatten, 2008), whereby seeing the target items in a diverse context each time would lead learners to process them for meaning each time as well, leading to longer processing times in the Different condition.

In terms of early measures, we found significant fixed effects of both condition and reading time on FFD, which is said to represent aspects of processing such as recognition and lexical access (Godfroid, 2019), although there was no significant interaction between the two. Our results indicate that encounters with novel items in the Different condition produced significantly longer initial fixations than in the Same condition, and that first fixation length decreased across the three readings overall. This appears consistent with the SDM (Semantic Distinctiveness Model, see Johns, Dye & Jones, 2014 and Jones, Johns & Recchia, 2012), which may suggest that items encountered in diverse conditions would require more time to be encoded, given that the different contexts would not coincide with existing memory traces, although at each subsequent re-encounter, as the representation grows stronger, the time for lexical access would decrease. However, as the interaction between reading time and condition for FFD did not reach statistical significance in our case, further investigation would be needed to confirm this interpretation. Future studies could also incorporate posttest processing measures, as in Pagan and Nation (2019), to further explore whether the different contextual conditions lead to processing differences that are also reflected in posttest processing tasks.

On exploring the effects of contextual diversity on incidental vocabulary learning, we found no significant effect of condition on participants' ability to

recognize the form of novel items. This may be because form recognition is one of the easiest components of vocabulary knowledge (Nation, 2001), implying that contextual factors may not play a significant role in accuracy of form recognition. Previous research, however, has shown faster and more accurate form recognition under contextually diverse conditions (e.g., Johns et al., 2016). This would also be predicted by the SDM (Johns, Dye & Jones, 2014; Jones, Johns & Recchia, 2012) which would indicate an advantage for the Different condition, although this was not borne out in our study. This might be due to the fact that we operationalized contextual diversity as document count, rather than using the measure of “semantic distinctiveness” as in Johns et al. (2016), or because we studied L2 learners with upper-intermediate proficiency, who may be expected to process items differently to L1 speakers. Future studies would thus be required to confirm or elaborate on the effects of contextual diversity on form recognition under different experimental conditions.

Our participants’ ability to recall the meaning of the target was overall very low which is why the results could not be analysed using inferential statistics or LMM. The small gains in terms of meaning recall are in line with other studies on vocabulary learning through reading (e.g., Brown et al., 2008; Pellicer-Sánchez & Schmitt, 2010). These results are expected considering that reading is a receptive task and given the difficulty in acquiring this dimension of vocabulary knowledge (Gonzalez-fernandez & Schmitt, 2020).

In terms of meaning recognition, gains were significantly higher in the immediate posttest for participants who encountered words in the Same contextual condition rather than across Different contexts. This is contrary to some existing research (e.g., Ferreira & Ellis, 2016; Liu & Todd, 2016). Ferreira and Ellis (2016) found that participants who had encountered obscure words in different sentences recognized their meaning faster and made fewer errors than those who had seen the targets in the same sentences repeated six times. However, since their study involved reading sentences rather than whole texts, as well as intentional learning (participants were told to try to infer the meaning of unknown words), the discrepancy in the results is not unexpected. Our results also contradict Liu and Todd’s (2016) study. We can speculate that the high number of repetitions of the same text (seven) in Liu and Todd (2016) probably contributed to a consequent loss of interest/attention and/or need to infer individual word meanings more clearly than in our study, which only included three repeated readings of the same text. This remains speculative as there was no processing information in Liu and Todd (2016).

In contrast, our findings regarding meaning recognition have been borne out by other studies that suggested that contextual uniformity may help build more stable semantic representations (e.g., Johns et al., 2016). This can be explained by the fact that a stable or repeated context would make it easier for participants to first discern and subsequently confirm their guesses of the meaning of novel words from contextual clues. The changes in these clues inherent in the different context might make it more difficult to accurately infer meanings across the texts.

Furthermore, as posited by other researchers, the positive effects of contextual diversity may be linked to higher proficiency levels (e.g., Horvath & Arunachalam, 2021; Zahar et al., 2001) and have often been found among L1 rather than L2 learners (e.g., Adelman et al., 2006; Bolger et al., 2008). The present study involved

not only longer texts but also incidental learning of completely unknown pseudowords. We might reasonably assume that for participants of an upper-intermediate level (average B2 as per the Quick Oxford Placement), the difficulty of the task may account for the fact that the Same condition was more conducive to initial meaning recognition since it would allow readers to become more familiar with the contextual clues through repetition (Liu & Todd, 2016), leading to a better chance of inferring meanings in that condition.

Notwithstanding these explanations, it is interesting to note that the difference between the two conditions is not present in the delayed posttest scores, indicating that while the Same condition may have led to immediately greater gains, it also led to a greater loss of knowledge in the delayed posttest. This seems to suggest that despite significantly lower immediate recognition, the Different condition appears to have fostered more stable long-term retention. This might be explained by the fact that contextual diversity allows for more context-independent encoding in the lexicon (Pagán & Nation, 2019) and easier future retrieval on the basis of “likely need” (Anderson & Milson, 1989; Anderson & Schooler, 1991). It could also be explained with reference to the SDM or semantic distinctiveness model (Johns & Jones, 2008; Johns, Dye & Jones, 2014; Jones, Johns & Recchia, 2012), which posits that contextual diversity aids encoding items in the lexical memory at a “stronger magnitude” (Jones *et al.*, 2017, p. 270) than contextual repetition, which might imply better long-term retention. It could also be interpreted in light of the desirable difficulty framework (cf. Bjork, 1994; Suzuki, *et al.*, 2019). In this case, the Different condition constituted a “difficulty,” reflected in participants’ slower online processing across readings, but such difficulty was “desirable,” as it fostered long-term retention of knowledge. This is in line with other “desirable difficulties” in L2 practice conditions that have been shown to be more challenging during the learning phase, but which foster long-term knowledge (Suzuki *et al.*, 2019).

Finally, our results regarding the relationship between processing and vocabulary learning suggest that processing, as analysed by TFD, does not differentially affect vocabulary learning under Same vs. Different repeated reading conditions, as the interaction between TFD and context was not significant. However, our results do suggest a link between processing and overall vocabulary gains. Existing literature on the relationship between processing and vocabulary learning is divided: some studies have found a positive relationship between longer processing times and greater gains (Godfroid *et al.*, 2013; Pellicer-Sánchez, 2016), while others have not (e.g., Elgort *et al.*, 2018; Pellicer-Sánchez *et al.*, 2022; Serrano & Pellicer-Sánchez, 2024), or have found inconclusive results (Wochna & Juhasz, 2013).

We found that longer TFD may be linked to higher scores in form recognition (irrespective of condition and test time). We also found a significant effect of processing on meaning recognition scores overall, with the effect being significantly greater at the immediate posttest than at the delayed posttest. These results are in line with previous studies that have found longer processing times to be related to greater gains (e.g., Koval, 2019; Mohamed, 2018; Pellicer-Sánchez, 2016). This finding can be explained by considering longer processing times as an indication of greater cognitive engagement (Godfroid *et al.*, 2013), which could be expected to lead to greater learning. The results are also coherent with existing theories that suggest noticing and attention to target items are key to learning (e.g., Robinson

1995; Schmidt, 2001). In terms of the interaction between test time and the effects of processing on meaning recognition scores, we could speculate that the benefits conferred by greater processing may diminish over time if there is no further exposure to the target words, thus leading to a much weaker effect at the delayed posttest. This would need to be tested and confirmed in future research.

Our study constitutes an initial exploration on how processing might predict different degrees of incidental vocabulary learning when new words are repeated in the same versus different texts. Future replication studies would be essential to reach more robust conclusions about the links between contextual diversity, processing, and long- and short-term vocabulary learning. Incorporating further vocabulary testing measures and using stimulated recalls or post-task questionnaires on engagement and learning strategies employed by participants might help better explain the links between processing patterns and vocabulary gains.

Limitations

This study was not without limitations, one of which may be the lack of ecological validity in the task, particularly under the Same condition. This is because learners are rarely asked to read a text three times consecutively in one sitting with general comprehension being the only focus. Future studies could perhaps devise other (non-vocabulary related) exercises to make this a more ecologically valid exercise for participants.

Another limitation is that, while our study focused on the role of contextual diversity, while trying to control for other potential factors, it is possible the role of contextual diversity in processing and learning of new words may be affected by such factors. Future studies should explore how the effects of contextual diversity may be affected by variables such as frequency, participant-, text-, and treatment-related variables including age, proficiency, engagement, modes of input, test format, and others (see Uchihara et al., 2019) that may play a key role in vocabulary processing and learning. These are aspects that would be worth investigating in conjunction with contextual diversity.

In addition, triangulating a wider set of variables with more online and offline measures (the latter including stimulated recalls or further vocabulary tests, for example) might allow us to better understand the subprocesses involved in vocabulary acquisition through reading (Godfroid & Winke, 2015; Pellicer-Sánchez, 2020) under different textual and experimental conditions.

Conclusion

This study explored the effects of contextual diversity versus sameness in a repeated reading design on the processing and incidental learning of novel vocabulary. Our findings suggest that while contextual sameness facilitates faster processing across subsequent readings and short-term learning of new words, it does not necessarily lead to better vocabulary learning long term. While initial encoding and lexical integration appear to take longer in the contextually Different condition, we did not find that these longer processing times were later reflected in higher vocabulary learning scores. This suggests that the role of context may depend on factors such as

how diversity is conceptualized, the types of contexts new words are presented in, whether the target language is in the L1 or L2, and the proficiency of the participants, among others. Future studies should investigate other types of learners (differing in age, proficiency, etc.) as well as other types of texts or test formats, in order to generalize these results.

Supplementary material. To view supplementary material for this article, please visit <https://doi.org/10.1017/S0142716424000407>

Acknowledgements. This research was funded by grants PID2019-110536GB-I00, from the Spanish Ministry of Science and Innovation, and 2023SGR00303, from the Catalan Agency for Management of University and Research Grants (AGAUR). We would like to thank the editor and the anonymous reviewers for their insightful comments.

Competing interests. The authors declare none.

Notes

1 L1 (TFD $p = .11$; AFD $p = .52$; FFD $p = .40$; FC $p = .15$).

2 In order to interpret the exponential coefficient better, the variable TFD, originally in milliseconds, was recoded in seconds.

References

- Adelman, J. S., Brown, G. E., & Quesada, J. G. M. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological Science*, *17*(9), 814–823. <https://doi.org/10.1111/j.1467-9280.2006.01787.x>
- Alotaibi, S., Pellicer-Sánchez, A., & Conklin, K. (2022). The effect of input modes and number of exposures on the learning of L2 binomials. *ITL – International Journal of Applied Linguistics*, *173*(1), 58–93. <https://doi.org/10.1075/itl.21001.alo>
- Anderson, J. R., & Milson, R. (1989). Human memory: an adaptive perspective. *Psychological Review*, *96*(4), 703–719. <https://doi.org/10.1037/0033-295X.96.4.703>
- Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science*, *2*(6), 396–408. <https://doi.org/10.1111/j.1467-9280.1991.tb00174>
- Beck, I. L., McKeown, M. G., & McCaslin, E. S. (1983). Vocabulary development: all contexts are not created equal. *The Elementary School Journal*, *83*(3), 177–181. <http://www.jstor.org/stable/1001501>
- Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). MIT Press.
- Bolger, D. J., Balass, M., Landen, E., & Perfetti, C. A. (2008). Context variation and definitions in learning the meanings of words: an instance-based learning approach. *Discourse Processes*, *45*(2), 122–159. <https://doi.org/10.1080/01638530701792826>
- Brown, R., Waring, R., & Donkaewbua, S. (2008). Incidental vocabulary acquisition from reading, reading-while-listening, and listening to stories. *Reading in a Foreign Language*, *20*(2), 136–163.
- Chen, C., & Truscott, J. (2010). The effects of repetition and L1 lexicalization on incidental vocabulary acquisition. *Applied Linguistics*, *31*(5), 693–713. <https://doi.org/10.1093/applin/amq031>
- Conklin, K., Pellicer-Sánchez, A., & Carrol, G. (2018). *Eye-tracking: A guide for Applied Linguistics Research*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781108233279>
- de Groot, A. M. B., & van Hell, J. G. (2005). The Learning of Foreign Language Vocabulary. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 9–29). New York: Oxford University Press. <https://doi.org/10.1093/oso/9780195151770.003.0002>
- Elgort, I., Brysbaert, M., Stevens, M., & Van Assche, E. (2018). Contextual word learning during reading in a second language. *Studies in Second Language Acquisition*, *40*(2), 341–366. <https://doi.org/10.1017/S0272263117000109>

- 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. <https://doi.org/10.1080/23273798.2014.942673>
- 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(2), 365–414. <https://doi.org/10.1111/lang.12052>
- Ferreira, R. L. R., & Ellis, A. D. (2016). Effects of contextual diversity on semantic decision and reading aloud: evidence from a word learning study in English as a second language/Efectos de la diversidad contextual en la decisión semántica y la lectura en voz alta: evidencia de aprendizaje de palabras en inglés como segundo idioma. *Estudios De Psicología*, 37(1), 1–21. <https://doi.org/10.1080/02109395.2015.1122433>
- Godfroid, A. (2019). *Eye tracking in second language acquisition and bilingualism A research synthesis and methodological guide*. Routledge. <https://doi.org/10.4324/9781315775616>
- Godfroid, A., Ahn, J., Choi, I., Ballard, L., Cui, Y., Johnston, S., Lee, S., Sarkar, A., & Yoon, H. J. (2018). Incidental vocabulary learning in a natural reading context: an eye-tracking study. *Bilingualism: Language and Cognition*, 21(3), 563–584. <https://doi.org/10.1017/s1366728917000219>
- 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(3), 483–517. <https://doi.org/10.1017/S0272263113000119>
- Godfroid, A., & Winke, P. (2015). Investigating implicit and explicit processing using L2 learners' eye-movement data. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 325–348). John Benjamins. <https://doi.org/10.1075/sibill.48.14god>
- Gonzalez-fernandez, B., & Schmitt, N. (2020). Word knowledge: exploring the relationships and order of acquisition of vocabulary knowledge components. *Applied Linguistics*, 41(4), 481–505. <https://doi.org/10.1093/applin/amy057>
- Horst, M. (2000). *Text encounters of the frequent kind: Learning L2 vocabulary through reading*. Unpublished doctoral dissertation, University of Wales, Swansea.
- Horst, M. (2005). Learning L2 vocabulary through extensive reading: a measurement study. *Canadian Modern Language Review*, 61(3), 355–382. <https://doi.org/10.3138/cmlr.61.3.355>
- Horst, M., Cobb, T., & Meara, P. (1998). Beyond a clockwork orange: acquiring second language vocabulary through reading. *Reading in a Foreign Language*, 11(2), 207–223.
- Horvath, S., & Arunachalam, S. (2021). Repetition versus variability in verb learning: sometimes less is more. *Journal of Speech Language and Hearing Research*, 64(11), 4235–4249. https://doi.org/10.1044/2021_jslhr-21-00091
- Hulstijn, J. (2001). Intentional and Incidental Second Language Vocabulary Learning: A Reappraisal of Elaboration, Rehearsal, and Automaticity. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 258–286). Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139524780.011>
- IBM Corp. Released 2020. IBM SPSS Statistics for Windows, Version 27.0. Armonk, NY: IBM Corp
- Johns, B. T., Dye, M., & Jones, M. P. (2016). The influence of contextual diversity on word learning. *Psychonomic Bulletin & Review*, 23(4), 1214–1220. <https://doi.org/10.3758/s13423-015-0980-7>
- Johns, B. T., Dye, M. W., & Jones, M. N. (2014). The influence of contextual variability on word learning. *Proceedings of the 35th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.
- Johns, B. T., & Jones, M. N. (2008). Predicting word-naming and lexical decision times from a semantic space model. In *Proceedings of the 30th Annual Cognitive Science Society*. <https://escholarship.org/uc/item/7s0927cq>
- Jones, M. P., Dye, M., & Johns, B. T. (2017). Context as an Organizing Principle of the Lexicon. In *Psychology of learning and motivation* (pp. 239–283). Elsevier BV. <https://doi.org/10.1016/bs.plm.2017.03.008>
- Jones, M. P., Johns, B. T., & Recchia, G. (2012). The role of semantic diversity in lexical organization. *Canadian Journal of Experimental Psychology*, 66(2), 115–124. <https://doi.org/10.1037/a0026727>
- Koval, N. G. (2019). Testing the deficient processing account of the spacing effect in second language vocabulary learning: Evidence from eye tracking. *Applied Psycholinguistics*, 40(05), 1103–1139. <https://doi.org/10.1017/s0142716419000158>

- Laufer, B. (2005). Focus on form in second language vocabulary learning. *EUROSLA Yearbook: Annual Conference of the European Second Language Association*, 5, 223–250. <https://doi.org/10.1075/eurosla.5.11lau>
- Liu, Y. T., & Todd, A. C. (2016). Implementation of assisted repeated reading techniques for the incidental acquisition of novel foreign vocabulary. *Language Teaching Research*, 20(1), 53–74. <https://doi.org/10.1177/1362168814559802>
- Meara, P., & Miralpeix, I. (2017). *Tools for researching vocabulary* (1st ed., Vol. 105). Channel View Publications. <https://doi.org/10.21832/9781783096473>
- Mohamed, A. S. (2018). Exposure frequency in L2 reading. *Studies in Second Language Acquisition*, 40(2), 269–293. <https://doi.org/10.1017/s0272263117000092>
- Nagy, W. E., Anderson, R. C., & Herman, P. A. (1987). Learning word meanings from context during normal reading. *American Educational Research Journal*, 24(2), 237–270. <https://doi.org/10.3102/00028312024002237>
- Nagy, W. E., Herman, P. A., & Anderson, R. C. (1985). Learning words from context. *Reading Research Quarterly*, 20(2), 233–253. <https://doi.org/10.2307/747758>
- Nassaji, H. (2006). The relationship between depth of vocabulary knowledge and L2 learners' lexical inferencing strategy use and success. *The Modern Language Journal (Boulder, Colo.)*, 90(3), 387–401. <https://doi.org/10.1111/j.1540-4781.2006.00431.x>
- Nation, I. S. P. (1990). *Teaching and learning vocabulary*. Boston, MA: Heinle & Heinle.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge, England: Cambridge University Press. <https://doi.org/10.1017/CBO9781139524759>
- Nation, I. S. P. (2006a). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review*, 63(1), 59–81. <https://doi.org/10.1353/cml.2006.0049>
- Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9781139858656>
- Pagán, A., & Nation, K. (2019). Learning words via reading: contextual diversity, spacing, and retrieval effects in adults. *Cognitive Science*, 43(1), e12705. <https://doi.org/10.1111/cogs.12705>
- Pellicer Sanchez, 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. <https://doi.org/10.1017/S014271642200011X>
- Pellicer-Sánchez, A. (2016). Incidental L2 vocabulary acquisition from and while reading: an eye-tracking study. *Studies in Second Language Acquisition*, 38(1), 97–130. <https://doi.org/10.1017/S0272263115000224>
- Pellicer-Sánchez, A. (2020). Expanding english vocabulary knowledge through reading: insights from eye-tracking studies. *RELC Journal*, 51(1), 134–146. <https://doi.org/10.1177/0033688220906904>
- Pellicer-Sanchez, A., & Schmitt, N. (2010). Incidental vocabulary acquisition from an authentic novel: Do things fall apart? *Reading in a Foreign Language*, 22(1), 31–55. <https://doi.org/10.125/66652>
- Peters, E. (2020). Factors affecting the learning of single-word items 1. In *The Routledge Handbook of Vocabulary Studies* (1st ed., pp. 125–142). Routledge. <https://doi.org/10.4324/9780429291586-9>
- Pigada, M., & Schmitt, N. (2006). Vocabulary acquisition from extensive reading: a case study. *Reading in a Foreign Language*, 18(1), 1–28. <https://doi.org/10.125/66611>
- Pulido, D. (2007). The relationship between text comprehension and second language incidental vocabulary acquisition: a matter of topic familiarity? *Language Learning*, 57(s1), 155–199. <https://doi.org/10.1111/j.1467-9922.2007.00415.x>
- Rastle, K., Harrington, J., & Coltheart, M. (2002). 358,534 nonwords: The ARC Nonword database. *Quarterly Journal of Experimental Psychology*, 55A, 1339–1362. <https://doi.org/10.1080/02724980244000099>
- Rayner, K. (1977). Visual-attention in reading - eye-movements Reflect cognitive-processes. *Memory & Cognition*, 5(4), 443–448. <https://doi.org/10.3758/BF03197383>
- Rayner, K. (2009). Eye movements in reading: models and data. *Journal of Eye Movement Research*, 2(5), 1–10. <https://doi.org/10.16910/jemr.2.5.2>
- 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(3): 219–237. https://doi.org/10.1207/S1532799XSSR0703_2
- Reynolds, B. L., & Wible, D. (2014). Frequency in incidental vocabulary acquisition research: an undefined concept and some consequences. *TESOL Quarterly*, 48(4), 843–861. <https://doi.org/10.1002/tesq.197>

- Robinson, P.** (1995). Attention, memory, and the “Noticing” hypothesis. *Language Learning*, *45*(2), 283–331. <https://doi.org/10.1111/j.1467-1770.1995.tb00441.x>
- Rogers, J., & Cheung, A.** (2021). Does it matter when you review? Input spacing, ecological validity, and the learning of L2 vocabulary. *Studies in Second Language Acquisition*, *43*(5), 1138–1156. <https://doi.org/10.1017/S0272263120000236>
- Rott, S.** (1999). The effect of exposure frequency on intermediate language learners’ incidental vocabulary acquisition through reading. *Studies in Second Language Acquisition*, *21*, 589–619. <https://doi.org/10.1017/S0272263199004039>
- Schmidt, R.** (2001). Attention. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp.3–32). Cambridge University Press. <http://doi.org/10.1017/CBO9781139524780.003>
- Schmitt, N.** (2008). Review article: instructed second language vocabulary learning. *Language Teaching Research*, *12*(3), 329–363. <https://doi.org/10.1177/1362168808089921>
- Schmitt, N.** (2010). *Researching vocabulary: a vocabulary research manual*. Palgrave Macmillan. <http://doi.org/10.1057/9780230293977>
- Schmitt, N., & Schmitt, D.** (2020). *Vocabulary in language teaching (Second edition)*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781108569057>
- Serrano, R., & Pellicer-Sánchez, A.** (2024). Online processing and vocabulary learning in massed versus spaced repeated reading. *Vigo International Journal of Applied Linguistics*, *21*, 129–164. <https://doi.org/10.35869/vial.v0i21.4402>
- Sonbul, S., & Schmitt, N.** (2010). Direct teaching of vocabulary after reading: is it worth the effort? *ELT Journal*, *64*(3), 253–260. <https://doi.org/10.1093/elt/ccp059>
- Suzuki, Y., Nakata, T., & DeKeyser, R.** (2019). The desirable difficulty framework as a theoretical foundation for optimizing and researching second language practice. *The Modern Language Journal (Boulder, Colo.)*, *103*(3), 713–720. <https://doi.org/10.1111/modl.12585>
- Tobii Pro AB** (2014). Tobii Pro Lab (Version 1.142) [Computer software]. Danderyd, Sweden: Tobii Pro AB.
- Uchihara, T., Webb, S., & Yanagisawa, A.** (2019). The effects of repetition on incidental vocabulary learning: a meta-analysis of correlational studies. *Language Learning*, *69*(3), 559–599. <https://doi.org/10.1111/lang.12343>
- VanPatten, B.** (2008). Processing matters in input enhancement. In: T. Piske, & M. Young-Scholten (Eds.), *Input matters* (pp. 47–61). Clevedon: Multilingual Matters. <https://doi.org/10.21832/9781847691118-005>
- Waring, R., & Takaki, M.** (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language*, *15*(2), 130–163. <https://doi.org/10.10125/66776>
- Webb, S.** (2008). The effects of context on incidental vocabulary learning. *Reading in a Foreign Language*, *20*(2), 232–245. <https://doi.org/10.10125/66826>
- Webb, S., & Chang, A. C.** (2012). Vocabulary learning through assisted and unassisted repeated reading. *Canadian Modern Language Review-revue Canadienne Des Langues Vivantes*, *68*(3), 267–290. <https://doi.org/10.3138/cmlr.1204.1>
- Webb, S., & Chang, A. C.-S.** (2015a). Second language vocabulary learning through extensive reading with audio support: how do frequency and distribution of occurrence affect learning? *Language Teaching Research*, *19*(6), 667–686. <https://journals.sagepub.com/doi/10.1177/1362168814559800>
- Webb, S., Yanagisawa, A., & Uchihara, T.** (2020). How effective are intentional vocabulary-learning activities? A meta-analysis. *The Modern Language Journal*, *104*(4), 715–738. <https://doi.org/10.1111/modl.12671>
- Webb, S. A.** (Ed.). (2020). *The Routledge handbook of vocabulary studies*. New York, New York. <https://doi.org/10.4324/9780429291586>
- Wochna, K. L., & Juhasz, B. J.** (2013). Context length and reading novel words: an eye-movement investigation. *British Journal of Psychology*, *104*(3), 347–363. <https://doi.org/10.1111/j.2044-8295.2012.02127.x>
- Zahar, R., Cobb, T., & Spada, N.** (2001). Acquiring vocabulary through reading: effects of frequency and contextual richness. *Canadian Modern Language Review-revue Canadienne Des Langues Vivantes*, *57*(4), 541–572. <https://doi.org/10.3138/cmlr.57.4.541>

Cite this article: Chandy, R., Serrano, R., & Pellicer-Sánchez, A. (2024). The effect of context on the processing and learning of novel L2 vocabulary while reading. *Applied Psycholinguistics* *45*, 1086–1113. <https://doi.org/10.1017/S0142716424000407>