Online processing and vocabulary learning in massed versus spaced repeated reading

Raquel Serrano
Dept. Modern Languages and Literatures and English Studies
University of Barcelona
raquelserrano@ub.edu

Ana Pellicer-Sánchez
Department of Culture, Communication & Media
UCL Institute of Education
University College London
a.pellicer-sanchez@ucl.ac.uk

Abstract

The present exploratory study analyzed the effect of two different implementations of repeated reading (massed vs. spaced) on the online processing of a new text including some unknown vocabulary, as well as on incidental vocabulary learning. The relationship between processing and learning of target vocabulary in massed versus spaced repeated reading was also examined. A group of Spanish undergraduate students read the same text in English three times under either a massed (three times consecutively) or spaced (once a week) schedule, while their eye movements were recorded with an eye tracker. Knowledge of a set of 12 target words from the text was assessed through multiple-choice meaning-recognition tests. The results of the statistical analyses examining processing of the whole text suggest that repeated reading led to a significant decrease in total reading time and number of fixations in the massed group but not in the spaced group. However, in the analyses at the word level, no significant differences were found between the two conditions in the processing of the target words or immediate vocabulary gains. Finally, it was observed that longer processing time on the target words was related to larger immediate vocabulary gains, but only in the spaced group.

Keywords: repeated reading, input spacing, online processing, vocabulary learning, eye-tracking.

Resumen

El presente estudio exploratorio analiza el efecto de dos implementaciones diferentes de lectura repetida (concentrada vs. espaciada) en el procesamiento del texto así como en el aprendizaje incidental de nuevo vocabulario. También se examina la relación entre procesamiento y aprendizaje. Un grupo de estudiantes universitarios españoles leyó el mismo texto en inglés tres veces de forma consecutiva y otro grupo de forma espaciada (una vez a la semana), mientras se registraban sus movimientos oculares. Una prueba de selección múltiple evaluó el reconocimiento del significado de 12 palabras meta incluidas en el texto. Los análisis estadísticos sobre el procesamiento del texto completo muestran que la lectura repetida condujo a una disminución significativa en el tiempo total de lectura y en el número de fijaciones en el caso de la lectura repetida concentrada, pero no en el de la espaciada. Sin embargo, en los análisis a nivel de la palabra no se encontraron diferencias significativas entre las dos condiciones de lectura repetida ni en el procesamiento ni en el aprendizaje a corto plazo de nuevo vocabulario. Finalmente, se observó que un mayor tiempo de procesamiento estaba relacionado con un mejor aprendizaje de vocabulario a corto plazo, pero solo en el grupo espaciado.

Palabras clave: lecturas repetidas, distribución temporal, procesamiento, aprendizaje de vocabulario, seguimiento ocular.

1. Introduction

Developing reading fluency in English as a foreign language is crucial in many educational contexts, especially those that use English as the medium of instruction, which are becoming increasingly popular worldwide, but also in regular EFL contexts. Reading fluency has been defined in many different ways; however, under most definitions it entails fast and efficient processing of a written text (Grabe et al., 2015, p. 75). The most common component of reading fluency is reading rate, usually operationalized as words per minute. However, beyond reading rate, changes in physical aspects of reading (e.g., fixations, fixation durations) are also important to examine the development of reading fluency (Nation, 2009). Eye-tracking has been increasingly used as an objective measurement of reading fluency, providing rich information about processing effort during reading.

Considering the importance of reading fluency, researchers and practitioners have explored different ways to help learners process written texts more efficiently. One of these instructional approaches is repeated reading, which involves reading the same text multiple times. Research has shown that repeated reading leads to faster reading rates, in both the L1 and the L2 (Chang & Millet, 2013; Therrien,

2004). Additionally, research in L1 reading using eye-tracking has shown that repeated reading has a facilitation effect in processing efficiency, as manifested through shorter fixations, fewer fixations, and longer saccades (Hyönä & Niemi, 1990; Raney & Rayner, 1995).

While the examination of eye movements can also inform about changes in L2 reading behavior and processing efficiency (Conklin, Pellicer-Sánchez, & Carrol, 2018), no previous study has explored how reading patterns change when rereading the same text multiple times. Eye-tracking research in L2 reading has mostly focused on processing of unknown vocabulary, rather than processing of whole texts. The evidence coming from this line of research suggests that the processing of repeated exposures to novel words becomes less effortful as repetitions increase (Elgort, Brysbaert, Stevens, & Van Assche, 2018; Godfroid et al., 2018; Pellicer-Sánchez, 2016) and that the amount of attention given to novel items seems to be related to vocabulary gains (e.g., Godfroid, Boers, & Housen, 2013; 2018; Pellicer-Sánchez, 2016). It remains to be known whether these patterns are also found in the context of repeated reading.

Apart from helping learners read faster and more efficiently, L2 repeated reading has also been found to foster incidental vocabulary learning (e.g., Liu & Todd, 2016). Considering that repeated exposure to novel words is necessary for incidental learning to take place (Nation, 2013; Uchihara, Webb, & Yanagisawa, 2019), reading the same text several times guarantees that learners will encounter the same words multiple times.

When analyzing any type of repeated L2 practice, it is important to consider inter-repetition spacing, as it could affect both learning processes and products (Koval, 2022; Suzuki, Nakata, & DeKeyser, 2019). Research in cognitive psychology as well as in SLA has shown that massed repetitions are processed more easily because of recency effects; however, this lack of processing effort can have negative consequences in terms of learning outcomes (Koval, 2019). Although there are some studies that have analyzed vocabulary learning through repeated reading over several sessions under differently spaced schedules (Serrano & Huang, 2018, 2023), no previous studies have examined how reading patterns change when L2 learners reread the same text repeatedly in one versus several sessions, or how reading patterns affect vocabulary learning. Exploring how different inter-repetition spacing affects reading speed and incidental vocabulary learning in the context of repeated reading can provide interesting insights on the role of input spacing in L2 practice. Additionally, findings from such research can also have relevant pedagogical implications concerning the implementation of repeated reading in the L2 class.

Considering these gaps, the goal of the present exploratory study is to throw more light on the practice of repeated reading in an L2 by examining both processing as well

as learning outcomes when inter-repetition spacing is massed as opposed to spaced. In terms of processing, the study will use eye-tracking to examine changes in online processing of a text as a whole as well as a set of novel words included in it. Regarding learning products, the study will examine how different spacing conditions (massed vs. spaced) affect the degree of incidental vocabulary learning through repeated reading.

2. Literature review

2.1. Repeated Reading for Fluency and Vocabulary Learning

Repeated reading, which involves reading short texts multiple times, is a technique used in schools to promote L1 reading fluency and comprehension, for children with or without reading disabilities (Therrien, 2004). Reading involves the interaction of diverse complex processes including word decoding and meaning comprehension at the word, sentence and discourse level. According to La Berge and Samuels (1974), and Samuels (2004), repetitions make decoding easier and also help create associations between words, which contributes towards reading fluency. Similarly, repeated reading allows for greater familiarity with the content of the text, which also leads to faster reading (Samuels, 2004).

In the L2 literature, the studies by Gorsuch and Taguchi (2008) and Chang and Millet (2013) show that repeated-reading interventions (five repetitions in one single session) with audio support helped Vietnamese and Taiwanese adult EFL learners significantly improve their reading fluency and comprehension. The faster reading rates attained through repeated reading show that the reading process is becoming less effortful for the readers, and, as Grabe and Stoller (2013) suggest, these faster reading rates might be a sign of automatization of reading processes (DeKeyser, 2007).

A few studies in the L1 context have used eye-tracking to examine the effect of repeated reading on online reading behavior. These studies have shown that repeated reading generates a facilitation effect that is reflected in fewer and shorter fixations, longer saccades (Hyönä & Niemi, 1990; Inhoff, Topolski, Vitu, & O'Regan, 1993; Raney & Rayner, 1995) and a reduction in the proportion of regressions (Schnitzer & Kowler, 2006), presumably because repeated reading decreases attention demands (Inhoff et al., 1993). This facilitation effect concerns both surface features related to visuographic information, as well as higher-level comprehension processes (Hyönä & Niemi, 1990). Crucially, no previous studies have examined the effect of repeated reading on L2 reading behavior.

Research also shows that repeated reading can promote incidental vocabulary learning (e.g., Han & Chen, 2010; Horst & Meara, 1999; Liu & Todd, 2016; Llanes

& Tragant, 2021; Webb & Chang, 2012). The number of repetitions examined in previous research varies from a minimum of two up to eight. The fact that repeated reading guarantees multiple exposures to novel words is probably one of the features that contributes to vocabulary learning in this type of reading practice. However, some authors have also raised concerns about the fact that repeated readings of the same text can have a detrimental effect on learners' attention and motivation (Nichols, Rupley, & Rasinski, 2008). Concerning L2 learners' attention while reading, eye-tracking research has shown that encountering the same words several times in a text leads to a decrease in reading times across repetitions (Godfroid et al., 2018; Elgort et al, 2018; Pellicer-Sánchez, 2016). Moreover, eye-tracking studies have also provided some evidence that longer processing times on target vocabulary are related to better performance in vocabulary tests (e.g., Godfroid, Boers, & Housen, 2013; Pellicer-Sánchez, 2016). However, these questions have not been investigated in the context of repeated reading. Exploring how reading times change across repeated exposures to novel words and how such processing is related to vocabulary gains through repeated readings of the same text can provide insights into the potential of repeated reading to promote incidental vocabulary learning. Additionally, empirical studies should be conducted to explore whether different implementations of repeated reading could reduce the effect of boredom or lack of attention. One alternative could be spacing the repetitions of same-text re-readings over several sessions instead of just one (Horst & Meara, 1999; Llanes & Tragant, 2021).

2.2. Spacing and Repeated Reading

The discussion of spacing effects is particularly relevant for the application of repeated reading, as for any type of L2 practice that involves repetition. Studies in cognitive psychology have provided evidence for the spacing effect, which suggests that including spacing between repetitions of target items is more effective for learning than massed sequences, in which repetitions appear immediately (Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006). In addition, when considering spaced learning, it has been suggested that longer lags are more beneficial than shorter lags, although this lag effect is not as uniform or robust as the spacing effect (Toppino & Gerbier, 2014). In fact, if the lags are too widely spaced, they might prevent retrieval of previous presentations from memory, which might eliminate the benefits of repeated practice (Thios & D'Agostino, 1976; Toppino & Bloom, 2002).

Most of the evidence for the spacing effect in cognitive psychology comes from verbal learning through paired-associates. Studies in SLA using a similar paradigm have also confirmed the spacing effect for vocabulary learning in one single session (Koval, 2022; Nakata, 2015; Nakata & Suzuki, 2019). There are different theories that try to account for the spacing effect, including the deficient processing, encoding

variability or study-phase retrieval theories (see Edmonds, Gerbier, Palasis, & Whyte, 2021; Serrano, 2012; Toppino & Gerbier, 2014). The desirable difficulties theory proposed by Bjork (Bjork, 1994; 2018) has received recent attention in the SLA literature (Rogers & Leow, 2020; Suzuki et al., 2019). This theory suggests that spacing makes processing more effortful, but this increased processing effort translates into better learning outcomes.

As several authors have emphasized (Suzuki et al., 2019; Koval, 2022), most SLA studies on the spacing and lag effects have focused on learning products rather than online processing during the learning phase. The few existing studies, however, confirm that processing is more effortful when repetitions are spaced. In the study by Koval (2019), a group of English L1 speakers was asked to infer the meaning of 24 Finnish words which appeared in consecutive sentences (massed) or in sentences that were spaced. The results of the eye-tracking measures showed longer reading times for spaced items, which were better learned than massed. Koval (2022) compared learning and processing of Finnish-English paired-associates repeated in massed, short-spaced and long-spaced sequences, and found that participants were significantly faster in retrieving the L1 form after seeing the L2 cue in massed rather than in spaced presentations. However, the results of the vocabulary posttests showed a significant advantage for spaced repetitions, which were processed more effortfully during training. While Koval provided evidence for the processing and learning of novel vocabulary through repeated exposures under massed vs. spaced conditions in one session, little is known about how learning and processing of vocabulary changes when repetitions occur in one session as opposed to several sessions, or when they occur in the context of repeated reading.

Although there are no studies focusing on how spacing in repeated reading affects reading processing or reading speed, there is some research in this direction in the area of oral fluency which provides some interesting insights on how fluency might be affected by inter-repetition spacing of L2 practice. In the case of Japanese EFL learners, Suzuki (2021) found that three blocked (massed) repetitions of an oral task (i.e., AAA, BBB, CCC) led to more fluent speech processing than interleaved practice, in which learners repeated the same tasks over three different days (ABC, ABC, ABC). This increased fluency achieved in the blocked condition also transferred to other tasks. Similarly, Bui, Ahmadian, and Hunter (2019) found that immediate repetitions of the same task led to significantly more fluent oral production than when repetitions were distributed over several days. These findings also suggest that processing under more concentrated schedules becomes less effortful (i.e., more fluent) than when repetitions are widely spaced.

To the authors' knowledge, there is only one study that has examined the role of spacing on incidental learning of vocabulary through repeated reading of the same

text. In the context of a Taiwanese high school, Serrano and Huang (2018) investigated how different schedules of repeated reading practice facilitated the learning of 36 English words. The study used a meaning recognition vocabulary test and found that the short-spaced schedule (repeated reading everyday over five consecutive days) encouraged significantly more learning at the end of the treatment than the long-spaced condition (repeated reading sessions over five weeks with a 7-day intersession interval, ISI). When analyzing long-term gains, however, the study found that there was a statistically significant loss for the short-spaced group, but not for the long-spaced group. This result supports the prediction of the desirable difficulty framework in the sense that more spacing might have created a desirable difficulty during the learning sessions, which was beneficial for long-term retention. However, this study does not offer any insights on how spacing affected reading fluency and reading processes.

The current paper draws on previous research on repeated reading and also on spacing/lag effects in order to learn about how repeated reading of the same text affects L2 reading behavior, L2 incidental vocabulary learning, as well as the relationship between the two, in massed versus spaced repeated reading. Crucially, despite the claims made about different spacing schedules leading to different amounts of online processing during the learning phase and the potential that this has to explain spacing effects (Koval, 2019, 2022), no previous studies have examined how spacing affects processing in the context of L2 repeated reading, which could have valuable theoretical and practical implications.

3. Research Questions

This study aims to address the gaps identified in the review of the literature and examine how L2 learners process repetitions of the same text and target vocabulary under a massed (subsequent repetitions) versus spaced (7-day ISI) distribution. Another aim of the study is to analyze vocabulary learning under the two spacing conditions, as well as how online processing of the target words relates to vocabulary learning. The following questions were addressed:

1. Are there any differences in the online processing of the text and target vocabulary between massed and spaced repeated reading, as examined by readers' eye movements?

2. Are there any differences in incidental vocabulary gains between massed and spaced repeated reading conditions?

3. Are vocabulary gains in massed and spaced repeated reading related to online processing of the target words?

Based on the existing literature on repeated reading as well as eye-tracking research, we hypothesize that repeated reading will facilitate the development of reading fluency, evidenced by a decrease in number and duration of fixations (Hyona & Niemi, 1990; Raney & Rayner, 1995), with a more rapid decrease under the massed condition (Koval, 2019), in line with findings on L2 fluency (Bui et al., 2019; Suzuki, 2021). This more significant decrease under the massed condition is expected to be reflected in lower vocabulary gains (Koval, 2019). These hypotheses are in line with the predictions of the desirable difficulties framework, according to which, more difficulties (i.e., more processing effort) during training would contribute to more solid learning gains (Suzuki et al., 2019). Finally, processing time on novel words is expected to be positively related to vocabulary gains, regardless of the spacing schedule (Godfroid et al., 2018; Pellicer-Sanchez, 2016).

4. Methods

4.1. Participants

The participants were 44 Spanish/Catalan bilingual undergraduate students from the same university taking English Studies (females = 34, mean age = 21.3). The participants were recruited from three different groups on the same course. On average, they had a vocabulary size of 6,937 words (SD = 1,050), according to the V_YesNo Test (Meara & Miralpeix, 2017), and advanced English proficiency (C1, Common European Framework of Reference, CEFR), as assessed by the Oxford Quick Placement Test (UCLES, 2001) (M = 48/60, SD = 5.6). All the participants obtained course credit for their voluntary participation. Half of the participants (n = 22) were randomly assigned to the massed condition and the other half (n = 22) to the spaced (7-day ISI). After discarding outliers, students missing one or more tests, and students whose eye movements were not accurately recorded, the final number of students whose data was included in the study were the following: massed = 18, spaced = 19 in the vocabulary analyses; and massed = 19, spaced = 19 in the analysis of eye movements.

There were no significant differences between the two participant groups in the Oxford Quick Placement Test (massed: M = 46.1, SD = 5.7; spaced: M = 48.5, SD = 5.7; t(35) = -1.09, p = .285, d = -0.37), or in the V_YesNo Test (massed: M = 6845.9, SD = 1010.8; spaced: M = 6973.0, SD = 1148.2; t(35) = -.357, p = .724, d = -0.17).

4.2. Instruments

4.2.1. Reading Passage

The text was a short narrative from the book *New Proficiency Reading* (Stephens 2006), which provides practice for the Cambridge Proficiency Exam (equivalent to C2 in the CEFR). The original passage was modified to ensure that the words in the text would not pose any difficulty to the participants. Considering their advanced proficiency and that their overall vocabulary size was 6,937 words on average, it seemed safe to assume that they would be familiar with the first 4,000 most frequent words in English. The modified text was 751 words long and knowledge of the first 4,000 most frequent words provided a lexical coverage of 98%. The participants were asked to read the text naturally in order to understand the main ideas. A set of 22 *True/False* comprehension questions, including general questions about main idea units, was devised in order to make sure the students were reading the text for meaning. The comprehension questions were in English and did not require processing of the target vocabulary. In general, the students answered these questions without difficulty and there were no significant differences between the two conditions (massed: M =17.3, SD = 1.9; spaced: M = 17.3, SD = 1.8; $t(35) = -.065$, $p = .949$, $d = 0.02$).

4.2.2. Vocabulary Tests

Twelve target words (six nouns and six adjectives), which were likely to be unknown by participants were selected from the reading passage. The initial selection was based on the researchers' evaluation of the difficulty of the words and confirmed with a small group of learners of similar proficiency to the participants in the study. The knowledge of these words was pretested using a *Yes/No* checklist vocabulary test containing these words and 48 distractors: 12 high frequency nouns, 12 high frequency adjectives, 6 low frequency nouns, 6 low frequency adjectives and 6 pseudo-words (see Appendix A for the list of words and Appendix B for the *Yes/No* checklist vocabulary test). The participants were asked to select the words for which they knew both the form and the meaning. While this format has been subject to criticism due to the possibility of the participants overestimating their knowledge and for lacking verification of knowledge (Pellicer-Sánchez & Schmitt, 2012), it allows for the measurement of a relatively large number of items in a short time. This format was adopted for practical reasons. Due to time constraints, the pretest had to take place on the same day as the first reading, and we wanted to prevent excessive focus on the target words through testing. The large number of distractors allowed us to avoid making the target items particularly salient before the treatment started. It is possible that, without having to demonstrate their knowledge of the selected items, participants overestimated their knowledge of some words. However, while conservative, this allowed us to control even for initial, partial

knowledge of the target items. The reliability of this pretest was acceptable: Cronbach alpha = .757.

In order to capture vocabulary learning after the repeated reading sessions (posttest), a multiple-choice test was created including the 12 target words exclusively. The multiple-choice format was preferred since it involves a demonstration of knowledge. A meaning-recognition test was adopted because it typically captures more gains than recall tests (Waring & Takaki, 2003) and, considering that the words only appeared three times (once in each repetition of the text), it would have been challenging for students to demonstrate productive knowledge. In the multiple-choice test, the students had to choose the best definition for each target word, out of four possible definitions. The non-target definitions (from the Macmillan online dictionary) referred to words from the same linguistic category as the target words and were related to the content of the story. There was an "I don't know" option to minimize guessing (see Appendix C for the multiple-choice test). All the words in the definitions in the multiple-choice test were either among the first 3,000 most frequent words or were cognates, which ensured that the language could be understood by the target participants. Cronbach's alpha = .703 shows that the reliability of this test was acceptable.

4.3. Procedure

The data collection for this study was part of a larger project that also included other tasks related to students' L2 writing skills. For the present study, all the participants completed a pretest, treatment (reading the same text three times), an immediate posttest (after the last reading), and two delayed posttests (one and five weeks after the immediate posttest). Two weeks before the first reading session, the students completed the V_YesNo vocabulary size test. Having information about the participants' vocabulary size prior to the study was crucial to ensure the appropriate modification of the reading material. The test was performed during class time in a computer lab and typically took around 10 minutes to complete.

The rest of the sessions took place individually in a computer lab. In the second session, the students in the massed group firstly did the target vocabulary pretest (5 minutes), followed by the Oxford Quick Placement Test (20 minutes). The students then read the passage three times on a computer while their eye movements were recorded with a Tobii T120 eye tracker (Tobii, www.tobii.com). This is a remote, desktop eye-tracker, with the camera and infrared light integrated within the monitor. It has a sampling rate of 120 Hz, a typical accuracy of 0.5° and 0.2° resolution. There was a distance of approximately 64 cm between the participants and the screen, as suggested by the eye tracker manual. The text was displayed over 10 pages in bold

Courier New 14 font. The sentences were double-spaced and there were 8-10 lines on each page. On average, there were 74 words per page (range: 60-90). Participants had to press the space bar on the computer's keyboard in order to progress from one page to the next and they could not go back to previous pages. A 5-point calibration was performed at the beginning of the experiment and before each repetition of the text. The participants read at their own pace, with each reading typically lasting 3-5 minutes. After the final reading, the students performed a pen-and-paper comprehension test (5 minutes) and a writing task related to the larger project (20 minutes). It was important to introduce a task between the final reading and the vocabulary test so that previous exposure to the target words was not so recent. Therefore, for both the massed and spaced groups, there was a lag of approximately 25 minutes between the end of the last reading and the immediate vocabulary test. Participants were asked to come back to the lab 7 and 35 days after the immediate posttest, which is when the two delayed posttests were administered, in line with other studies on time distribution of L2 practice (Suzuki & DeKeyser, 2017). Participants were not told that they would be tested again on the same vocabulary and were expecting other tasks related to the larger project. The first retention interval (RI) was expected to be advantageous for the massed condition while the second was hypothesized to be more beneficial for the spaced (Rohrer & Pashler, 2007).

The procedure for the students in the spaced group was the same, with the exception that, instead of reading the passage three times in the second session, they only read it once and came back for two more sessions (with a 7-day ISI between them). Additionally, the learners in the spaced group were asked one general comprehension question orally after readings 1 and 2 to give some purpose to the reading activity, as they did not receive the comprehension test until they finished reading 3. Altogether, the data for the learners in the massed group was collected over four sessions and for the spaced group over six sessions (see Figure 1 for a summary of the procedure).

Figure 1: Procedure followed for the massed and spaced repeated reading groups

MASSED	ISI Days	**SPACED**	ISI Days
SESSION 1 V_YesNo Test		**SESSION 1** V_YesNo Test	
	15		15

SESSION 2 Pretest Oxford Quick Placement test Reading 1 Reading 2 Reading 3 Comprehension questions Writing task Immediate posttest	
	7
SESSION 3 7-day RI posttest	
	28
SESSION 4 35-day RI posttest	

SESSION 2 Pretest Oxford Quick Placement test Reading 1 1 general comprehension question	
	7
SESSION 3 Reading 2 1 general comprehension question	
	7
SESSION 4 Reading 3 Comprehension questions Writing task Immediate posttest	
	7
SESSION 5 7-day RI posttest	
	28
SESSION 6 35-day RI posttest	

5. Analyses

Eye-tracking measures of two different areas of interest were examined: the whole text and each of the target words. Eye-tracking studies on repeated reading in the L1 context have examined global eye-tracking measures based on entire sentences (e.g., Hyönä & Niemi, 1990) and entire passages (Raney & Rayner, 1995). In line with Raney and Rayner (1995), we examined measures at both the passage and the target word levels. By looking at both eye movements to the text and to specific target words, we were able to examine the effect that spacing had on reading the specific target items, as well as on overall reading behavior. Three eye-movement measures were examined: total reading time, fixation count, and average fixation duration. Total reading time and fixation count are late eye-movement measures that reflect lexical integration and are affected not only by lexical factors, but also by contextual, syntactic, and discourse-level properties of what is being read (Conklin et al., 2018). Average fixation duration was examined as it has been claimed to be useful when exploring how eye movements

unfold over time (Conklin et al., 2018). At the text level, it was expected that repeated reading would increase text familiarity, which would be reflected in shorter total reading time, shorter fixation durations, and fewer fixations on the text (Hyönä & Niemi, 1990; Raney & Rayner, 1995). Average fixation duration is sensitive to text difficulty and, therefore, the reduced difficulty of a repeated text might be reflected in shorter average fixation durations (Hyönä & Niemi, 1990). Similarly, at the target word level, it was expected that repeated encounters with the text would lead to fewer and shorter fixations on the target words (Godfroid et al., 2013, 2018; Pellicer-Sánchez, 2016).

In order to analyze vocabulary gains under the two conditions, we computed relative gains by item, applying a formula that has often been used in vocabulary research (e.g., Peters & Webb, 2018): relative gains = (number of learned words/ (number of target items – number of known words)) x 100. "Learned words" are those that were incorrect in the pretest and correct in the posttest; and "known words" are those that were correct both in pretest and posttest. The above formula can correct possible overestimations of students' knowledge in the *Yes/No* pretest, since it considers as "known" only the words that were correct at both testing times. We computed immediate relative gains (with pretest and posttest scores), RI-7 gains scores (with scores in the pretest and in the 7-day RI delayed posttest) and RI-35 gains (with scores in the pretest and in the 35-day RI delayed posttest).

The SPSS 27 program (IBM, 2020) was used to perform the statistical analyses. In order to examine eye movements (RQ1), a series of linear mixed models (LMM) with repeated measures were performed: condition (massed vs. spaced), reading time (1, 2, 3) and their interaction were included as fixed effects. Estimated marginal means were obtained for each factor, and p-values of pairwise comparisons were corrected using Bonferroni's method. The residuals were saved and analyzed for normality of distribution. The Kolmogorov-Smirnov test confirmed the normality of the residuals in all cases ($p > .05$) except for total reading time and fixation count for the target words. In that case, a logarithmic transformation was performed, after which the distribution of the residuals was normal ($p = .200$). In order to examine vocabulary gains (RQ2), we firstly performed a preliminary analysis to compare the two conditions at the pretest level. The t-test revealed that the massed group knew significantly fewer words ($M = 2.22$, $SD = 1.73$) than the spaced ($M = 3.48$, $SD = 1.54$); $t(35) = -2.32$, $p = .026$, $d = -0.76$. Because of this difference, the pretest scores were always added in the statistical models that were used to examine vocabulary gains. Another LMM with repeated measures was performed to examine vocabulary gains, with condition (massed vs. spaced), time (immediate, RI-7 and RI-35) and pretest scores as fixed effects. The interaction between condition and time was also explored. Finally, in order to address the third question, partial correlations were performed between the eye-tracking measures on the target

words and the vocabulary gains. Effect sizes were manually calculated using Cohen's d. The SDs for the estimated means were calculated by using the standard error (SE) applying the following formula: SD = sqrt(n) * SE. Effect sizes were interpreted using the benchmarks proposed by Plonsky and Oswald (2014).

6. Results

6.1. *Reading Patterns in Massed vs. Spaced Repeated Reading*

Reading patterns for the text were analyzed first. The descriptive statistics of these text measures at each reading time appear in Table 1.

Table 1: Descriptive statistics for eye-movement measures to the text across the three repeated readings (T1, T2, T3) by condition. Mean values and standard deviations (in parentheses) calculated per page of text. Fixation durations in milliseconds

	T1		T2		T3	
	Massed (n = 19)	Spaced (n = 19)	Massed (n = 19)	Massed (n = 19)	Massed (n = 19)	Massed (n = 19)
Total Reading Time	200,588 (53,770)	169,228 (61,224)	168,453 (40,323)	161,727 (55,237)	154,106 (40,089)	160,642 (48,003)
Fixation Count	97.74 (23.67)	84.16 (25.40)	84.35 (16.94)	83.82 (26.69)	77.59 (17.39)	80.79 (18.80)
Average Fix. Duration	201.5 (28.8)	193.8 (26.4)	197.7 (26.0)	186.8 (19.8)	194.3 (27.9)	191.5 (26.0)

The results of the LMMs are presented in Table 2. In the case of **total reading time**, the effect of condition was not significant (p = .451, d = 0.17), while there was a main effect of time: significant decrease time 1-2 (p = .019, d = 0.28), and time 1-3 (p = .001, d = 0.39). There was a significant condition*time interaction (p = .038). When examining each condition separately, it was found that the massed group significantly decreased their total reading time between time 1-2 (p = .007, d = 0.45) and 1-3 (p = .001, d = 0.65), although the effect sizes were small. No significant changes were registered between time 2-3. The total reading time on the text in the case of the spaced group did not significantly change across time.

The results for **fixation count** showed a similar pattern: no effect of condition (p = .552, d = 0.14), and significant effect of time (p = .002) and condition*time interaction (p = .027). Under the massed condition, participants decreased the number of fixations significantly between time 1 and 2 (p = .009, d = 0.44), and time 1 and

3 (p <.001, d = 0.65), and, in both cases, the effect size was small. The spaced group showed no significant changes. The results for **average fixation duration** showed no significant effect of condition (p = .343), time (p = .185), or the interaction between the two (p = .386).

Table 2: Results of the LMMs with eye-tracking measures to the text

Target	Source	F	df1	df2	p	Condition	Time	
TRT	Condition	.572	1	107	.451	d = 0.17		
	Time	7.35	2	107	.001		1-2: 2.63, p = .019, d = 0.28 2-3: 3.72, p = .296, d = 0.11 1-3: 3.72, p = .001, d = 0.39	
	Condition *Time	3.37	2	107	.038	1: 1.91, p = .058, d =0.44 2: .421, p = .674, d =0.10 3: -.399, p = .691, d=-0.09	Massed	Spaced
							1-2: 2.99, p = .007, d =0.45 2-3: 1.37, p = .174, d =0.20 1-3: 4.44, p < .001. d =0.65	1-2: .717, p = 1.00, d =0.10 2-3: .104, p = 1.00, d =0.02 1-3: .821, p = 1.00, d =0.12
FC	Condition	.356	1	107	.552	d = 0.14		
	Time	6.74	2	107	.002		1-2: 2.12, p = .072, d = 0.22 2-3: 1.50, p = .137, d = 0.16 1-3: 3.65, p = .001, d = 0.37	
	Condition *Time	3.74	2	107	.027	1: 1.91, p = .059, d = 0.44 2: .066, p = .947, d = 0.02 3: .450, p = .653, d = -0.10	Massed	Spaced
							1-2: 2.90, p = .009, d = 0.44 2-3: 144, p = .151, d = 0.22 1-3: 4.42, p < .001, d = 0.65	1-2: .076, p = 1.00, d = 0.01 2-3: .665, p = 1.00, d = 0.10 1-3: .741, p = 1.00, d = 0.11
AFD	Condition	.908	1	107	.343	d = 0.24		
	Time	1.71	2	107	.185			
	Condition *Time	.961	2	107	.386			

TRT = Total reading time; FC = Fixation count; AFD = Average fixation duration

Eye movements to the target words were then analyzed. The descriptive statistics appear in Table 3. The results were similar to those reported for the whole text: a decrease in reading times across time, with a more marked decrease in the massed group (see Figure 2 for a visual representation of results).

Table 3: Descriptive statistics for eye-tracking measures to the target words across the three repeated readings (T1, T2, T3) by condition. Mean values and standard deviations (in parentheses) calculated per target item. Fixation durations in milliseconds

	T1		T2		T3	
	Massed (n = 19)	Spaced (n = 19)	Massed (n = 19)	Massed (n = 19)	Massed (n = 19)	Massed (n = 19)
TRT	789.0 (433.3)	646.1 (269.30)	522.8 (202.3)	570.9 (342.77)	483.3 (149.58)	477.3 (163.83)
TRT Log transform	-.155 (.223)	-.235 (.222)	-.318 (.195)	-.305 (235)	-.335 (.136)	-.346 (.156)
TRT Log transform	3.35 (1.14)	2.87 (1.06)	2.31 (0.63)	2.72 (1.53)	2.17 (0.55)	2.30 (0.58)
FC Log transform	.501 (.151)	.429 (.169)	.345 (.138)	.386 (.205)	.325 (.112)	.346 (.130)
AFD	207.6 (59.12)	208.8 (64.9)	201.5 (59.0)	188.8 (55.1)	199.6 (50.9)	185.5 (42.9)

TRT = Total reading time; FC = Fixation count; AFD = Average fixation duration

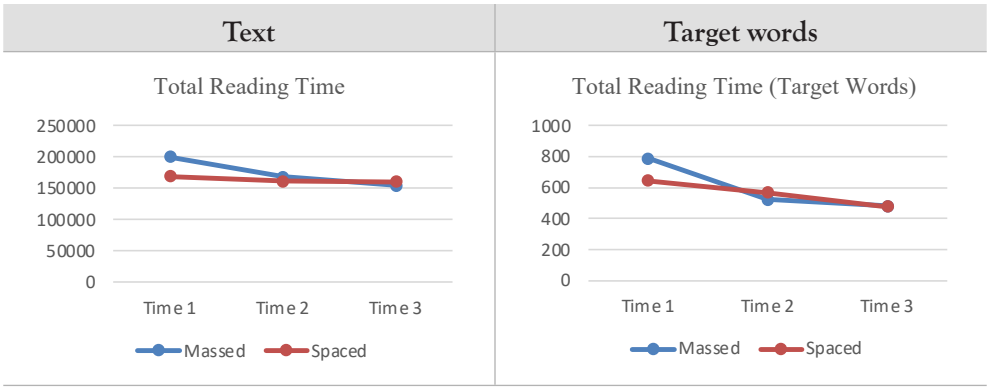
The results of the LMMs (see full report in Table 4) show that, for **total reading time**, there was a main effect of time ($p < .001$), with fixation durations significantly decreasing between time 1-2 ($p = .002$, $d = 0.43$), and time 1-3 ($p < .001$, $d = 0.52$). The effects of condition ($p = .608$, $d = 0.12$) and condition*time interaction ($p = .357$) were not significant. Similar results were found for **fixation count** to the target words, with no main effect of condition ($p = .910$, $d = 0.03$), no significant condition*time interaction ($p = .094$), and a significant main effect of time ($p < .001$), with the number of fixations decreasing significantly between time 1-2 ($p = .001$, $d = 0.45$) and time 1-3 ($p < .001$, $d = 0.60$). Finally, with respect to **average fixation duration** to the target words, the results were similar to those of the same measure for the whole text: no significant effect of condition ($p = .584$), time ($p = .174$), or condition*time interaction ($p = .637$).

Table 4: Results of LMM with eye-tracking measures to the target words

Target	Source	F	df1	df2	p	Condition	Time
TRT	Condition	.265	1	107	.608	d =0.12	
	Time	10.56	2	107	<.001		1-2: 3.44, p = .002, d =0.43 2-3: .868, p = .387, d =0.10 1-3: 4.35, p <.001, d =0.52
	Condition* Time	1.03	2	107	.357		
FC	Condition	.013	1	107	.910	d = 0.03	
	Time	12.67	2	107	<.001		1-2: 3.60, p = .001, d =0.45 2-3: 1.19, p = .236, d =0.15 1-3: 4.84, p < .001, d =0.60
	Condition* Time	2.42	2	107	.094		
AFD	Condition	.301	1	107	.584	d = 0.03	
	Time	1.78	2	107	.174		
	Conditio* Time	.453	1	107	.637		

TRT = Total reading time; FC = Fixation count; AFD = Average fixation duration

Figure 2: Results of the eye-movement measures to the text and target words



6.2. Relative Vocabulary Gains in Massed versus Spaced Repeated Reading

Table 5 shows the mean percentage of immediate and delayed relative vocabulary gains for each condition after controlling for pretest scores. The descriptive statistics show that the spaced group made more gains than the massed group, especially in the long term.

Table 5: Descriptive statistics vocabulary gains (in percentages) by condition and testing times, including estimated means and standard error in parentheses

Testing time	Experimental group	
	Massed (n = 18)	Spaced (n = 19)
Immediate gains (Pretest-posttest)	19.83 (4.17)	23.28 (4.05)
Gains RI-7 (Pre-delayed 1)	20.66 (4.24)	40.16 (4.09)
Gains RI-35 (Pre-delayed 2)	22.73 (4.16)	40.95 (3.98)

Continuous predictors are fixed at the following values: Pretest=2.934

The results of the LMM (see Table 6 for full report) showed that there were significant main effects of condition (p = .008, d = 0.65) and time (p <.001) with small effect sizes. The condition*time interaction was also significant (p = .006). When

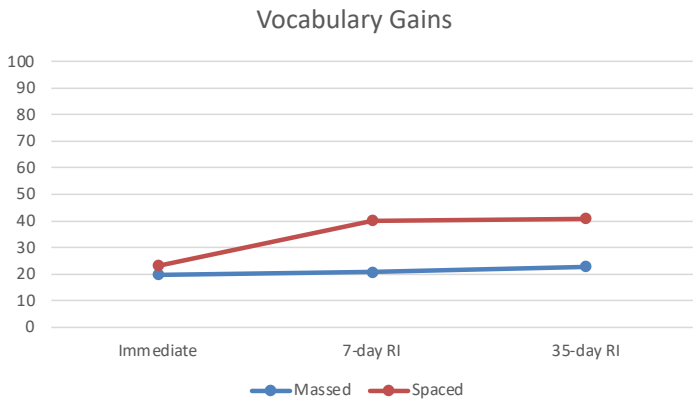
comparing conditions for immediate gains, no differences were observed between the two groups (p = .516, d = -0.16), while for long-term gains (both RI-7 and RI-35 gains), the spaced group significantly outperformed the massed group (p = .002, d = -0.78; p = .003, d = -0.73 respectively) and the effect size of this difference was medium. When examining the two conditions separately, the performance of the students in the massed group did not change significantly after the immediate posttest. The performance of the spaced group, on the other hand, significantly improved after the immediate posttest: immediate versus 7-day RI gains (p < .001, d = -0.69); immediate versus 35-day RI gains (p <.001, d = 0.73). Figure 3 illustrates how the performance differed between the two groups.

Table 6: Results of the LMMs with vocabulary gains

Source	F	df1	df2	p	Condition	Time	
Condition	7.25	1	115	.008	d = 0.17		
Time	9.45	2	115	< .001		1-2: -3.41, p = .002, d = -0.37 2-3: -5.39, p = .591, d = -0.06 1-3: -4.02, p < .001, d = -0.44	
Condition* Time	5.41	2	115	.006	1: -652, p = .516, d = -0.16 2: -3.22, p = .002, d = -0.78 3: -3.07, p = .003, d = -0.73	Massed	Spaced
						1-2: -.335 p = 1.00, d = -0.05 2-3: -.542 p = 1.00, d = -0.08 1-3: -.892 p = 1.00, d = -0.13	1-2: -4.53, p <.001, d = -0.69 2-3: -.215, p = .830, d = -0.03 1-3: -4.86, p <.001, d = -0.73

TRT = Total reading time; FC = Fixation count; AFD = Average fixation duration

Figure 3: Vocabulary gains (%) in massed and spaced repeated reading



These results suggest that the long-term gains experienced by the spaced group were probably not a direct product of the treatment, since more significant improvement took place after the treatment than during it.

6.3. *Relationship between Vocabulary Gains and Online Processing*

Next, we examined whether online processing of the target words was differentially related to vocabulary gains in the two conditions through partial correlations, controlling for pretest vocabulary scores. The results of the correlations indicate that cumulative reading times (each including time 1 + time 2 + time 3) were not related to vocabulary gains from repeated reading under the massed condition. Interestingly, the correlations were significant under the spaced condition, but only in the case of immediate vocabulary gains: total reading time (r = .582, p = .01), fixation count (r = .466, p = .05) and average fixation duration (r = .600, p = .009) (see Table 7).

Table 7: Results of the correlations between eye-tracking measures (T1 + T2 + T3) on the target words and vocabulary gains

		Adjusted Immediate			Adjusted RI-7			Adjusted RI-35		
		All (n = 37)	Massed (n = 18)	Spaced (n = 19)	All (n = 37)	Massed (n = 18)	Spaced (n = 19)	All (n = 37)	Massed (n = 18)	Spaced (n = 19)
TRT	**Corr.**	.257	-.139	.582	.097	.060	.145	.154	.025	.267
	Sig.	.131	.594	.01	.573	.819	.566	.370	.923	.283
FC	**Corr.**	.268	-.283	.466	.320	-.099	.100	.306	-.120	.184
	Sig.	.114	.271	.05	.057	.706	.692	.070	.646	.465
AFD	**Corr.**	.176	-.179	.600	.049	.139	.388	.90	.175	.393
	Sig.	.304	.492	.009	.776	.594	.111	.601	.502	.106

TRT = Total fixation duration; FC = Fixation count; AFD = Average fixation duration

7. Discussion

In this exploratory study, we analyzed the development of L2 reading fluency through two different implementations of repeated reading, examining online processing of text and target words through an eye-tracker. Another goal of the study was to analyze incidental vocabulary learning in massed versus spaced repeated reading as well as its relationship with online reading processing of the target words.

In response to RQ1, while the effect of the spacing condition on overall textual processing was not significant, a significant interaction was observed between time and spacing condition for two out of the three measures examined, namely total reading time and fixation count. Only the participants under the massed condition experienced a significant decrease across repetitions after reading the text the first time, which is in line with the facilitation effect found for repeated reading in the

L1 in previous eye-tracking research (Hyönä & Niemi, 1990; Raney & Rayner, 1995; Inhoff et al., 1993). This finding is also consistent with the facilitation effect of L2 repeated reading observed through reading rates (Chang & Millet, 2013; Gorsuch & Taguchi, 2008). In all these studies, repeated reading was done in one session, as was the case for the massed group in the present study. It can be assumed that faster reading rates and shorter processing times across repetitions happen because of the high activation of previous presentations of the same words and structures in the text. This increased familiarity with the materials makes the text easier to read and leads to a decrease in attentional demands. Efficient reading processing depends on the proceduralization and later automatization of lower-level processing skills, such as word recognition, syntactic parsing or propositional encoding (Grabe, 2009). High activation of the representations obtained from previous readings, as well as increased familiarization with the content of the text, made it easier for the learners under the massed condition to attain faster processing. This more efficient processing might be indicative of proceduralization, which, according to some studies, benefits from short-spaced/massed repetitions (Li & DeKeyser, 2019; Suzuki, 2020). It must be emphasized, however, that the effect sizes for the differences in processing were small, which may be due to the low number of repetitions, in comparison with other repeated reading studies (e.g., Chang & Millet, 2013; Gorsuch & Taguchi, 2008).

On the other hand, the fact that the students in the spaced group did not show differences in processing patterns across repetitions may suggest that they were processing the text each time as if it were a new text. A 7-day lag might have been a long interval for learners to be able to remember previous presentations of the text. In fact, study-phase retrieval theories of the spacing effect suggest that spacing is beneficial for learning as long as it allows for retrieval of previous presentations (Toppino & Bloom, 2002). Under the desirable difficulties framework, it can be said that, in this case, spacing added a difficulty that was not desirable for the development of automatic reading processes. This finding is also in line with previous findings for speech production, which show that more concentrated repetitions of the same task led to increased oral fluency in comparison to spaced repetitions (Bui et al., 2019; Suzuki, 2021).

The lack of interaction between reading time and condition for the eye-tracking measures on the target words suggests that there were no significant differences between the conditions with respect to the processing of the target vocabulary, despite the fact that the decrease in reading times is clearer for the massed than for the spaced condition. In this case, even though the learners under the massed condition had more recent exposure to previous presentations of the target words and their orthographic representations were more active, processing these words was probably still challenging because their meaning was unknown and probably difficult to guess.

This result is not in line with Koval's (2019), who found that target words in massed sequences were given less processing time than those in spaced sequences. This conflicting finding may be explained by methodological differences between the studies. The time between repetitions of the target items under the massed condition was shorter in Koval's research than in the current study, as the target words were introduced in short sentences in the participants' L1 that appeared subsequently, as opposed to a long text in the L2. Additionally, Koval's study examined intentional learning conditions, which might have affected how learners approached the reading task.

In relation to RQ2, the findings reported in this paper suggest that massed and spaced repeated reading did not lead to significantly different immediate vocabulary gains (19% and 23%, respectively). These gains are statistically significant, in line with other studies that have examined vocabulary learning through repeated reading in one session (e.g., Liu & Todd, 2016) or in more than one session (e.g., Llanes & Tragant, 2021). The results concerning long-term gains cannot be discussed in relation to the treatment, as intentional vocabulary learning possibly happened for some students in the spaced group, who improved their vocabulary scores after the treatment more significantly than during the treatment.

The fact that no significant differences were observed for incidental vocabulary learning between the two conditions at the immediate testing stage (short RI) is in line with other studies of distributed-practice effects for intentional learning both in cognitive psychology (Pavlik & Anderson, 2005) and SLA (Bird, 2010) which, despite showing a significant advantage for spaced conditions in the long term, failed to show an equivalent advantage at shorter RIs. According to Pavlik and Anderson (2005), at short RIs, the increased activation of massed items due to recency can make it possible for these items to be recalled as well as or even better than spaced items.

Our results are different from the studies by Koval (2019), and Nakata and Elgort (2021), which showed that spaced exposure significantly promoted more vocabulary gains than massed. This divergence in results can be explained by the fact that these two studies examined within-session spacing, whereas the current one investigated between-session spacing. Another reason that can explain the lack of differences is that incidental learning of vocabulary through just three repeated exposures to the target words was too demanding for all the students, regardless of the spacing of repetitions. While the immediate vocabulary gains registered as a product of repeated reading were in line with other previous studies on incidental vocabulary learning from reading (see Pellicer-Sánchez, 2016), they were quite low (about 20% of the words).

Finally, with respect to RQ3, about the relationship between processing and vocabulary learning, our results suggest that cumulative reading times and number of

fixations to the target words were related to immediate incidental vocabulary gains but only for the spaced condition. The fact that the correlations became non-significant for long-term gains in this condition might again point towards the fact that the scores in the delayed posttest were not a direct product of the treatment.

Previous eye-tracking studies have shown a positive relationship between reading times and vocabulary gains (Godfroid et al., 2013; Koval, 2019; Pellicer-Sánchez, 2016), suggesting that longer reading times might be evidence of more attention to language, or "noticing" of target features. Our results show that the students under the spaced condition who paid more attention to the target items were able to make more immediate vocabulary gains. This finding suggests that spacing might have imposed a desirable difficulty for those students who noticed the target words and devoted more time to processing them than for others who might have been more focused on understanding the meaning of the whole text. According to these results, it was especially important for readers under that condition, which was the most challenging, to spend time processing the target words in order for them to be able to identify their meaning later on.

8. Limitations

As explained in the methodology, for practical reasons, different vocabulary tests were performed for the pretest and posttest. However, the less challenging test was used as the pretest and, thus, a conservative approach was followed in the calculation of gains. Any effect that this may have would be expected to be similar under the two conditions. Another limitation is that the results of long-term learning could not be analyzed as a direct product of the treatment due to the significant gains experienced by the spaced group after the treatment, which were likely due to intentional learning and not repeated reading. Although this might be an interesting, positive consequence of spaced repeated reading, which promoted more engagement with vocabulary learning after the end of the treatment for that group (with interesting pedagogical implications), the analysis of this phenomenon is beyond the scope of this study. Future research could address this issue by performing between-participant testing, in which participants only complete one of the two delayed posttests. Alternatively, using pseudowords would also prevent intentional learning outside the treatment. Despite these limitations, the present study constitutes a valuable contribution to the literature, as it is one of the first to examine processing differences in massed versus spaced reading and, to the best of the authors' knowledge, the first to analyze processing differences when learners are only asked to read for meaning comprehension and not vocabulary learning. More studies should be conducted on this topic with more participants in different contexts in order to confirm the findings from this initial exploratory study.

9. Conclusion and Pedagogical Implications

The present study showed that spacing in repeated reading led to interesting processing differences, with participants under the massed condition experiencing a more significant decrease in number and duration of fixations to the text than those under the spaced condition. Target words were processed in a similar way regardless of the condition, indicating that the processing fluency created by being repeatedly exposed to novel items in reading seems to hold when the repetitions are differently spaced. This also indicates that attentional demands on novel words were the same regardless of the spacing schedule.

One possible pedagogical implication is that, when the goal of repeated reading is to develop faster reading speed, it is probably better to do massed repeated reading, as is normally done in repeated reading techniques implemented in schools to promote L1 literacy but also L2 reading fluency.

Furthermore, our results suggest that, unlike some previous findings, the time distribution of repeated reading episodes does not have any effect on subsequent short-term incidental vocabulary learning, when comparing massed and spaced exposures. The pedagogical implication that can be derived from this finding is that it is not so important whether repeated reading happens in one session or over several sessions if one of the goals is incidental vocabulary learning, but, considering the benefit of massed repetitions for fluency, teachers might be advised to promote repeated readings of the same text in one rather than over several sessions. However, more research should be performed before this recommendation can be made. First, it is only immediate gains that were analyzed in this study, and L2 teaching should be concerned with more durable gains, which could not be successfully examined in the present paper. Second, depending on the number of repetitions, massed practice can become counter-productive (Suzuki & Hanzawa, 2022).

Our findings also suggest a relationship between processing time on target words and vocabulary gains under the spaced condition, supporting the use of teaching techniques that attract learners' attention to target vocabulary, such as input enhancement (e.g., Barcroft, 2003; Kim, 2006), especially when readers are not likely to encounter repetitions of novel words within a short time span, and repeated reading happens on different days.

Finally, our study is one of the few that have examined how spacing affects L2 processing under repeated reading conditions; therefore, the results are not intended to provide conclusive evidence on the topic. To obtain more generalizable findings, more research should be conducted in this direction in order to understand the effect of spacing on reading processes and vocabulary learning through repeated reading.

Acknowledgments

This work was supported by Grant PID2019-110536GB-I00 of the Spanish Science Ministry.

References

Barcroft, J. (2003). Distinctiveness and bidirectional effects in input enhancement for vocabulary learning. *Applied Language Learning, 13*(2), 47–73.

Bird, S. (2010). Effects of distributed practice on the acquisition of second language English syntax. *Applied Psycholinguistics, 31*(4), 635–650. https://doi.org/10.1017/S0142716410000172

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). Cambridge, MA: MIT Press.

Bjork, R. A. (2018). Being suspicious of the sense of ease and undeterred by the sense of difficulty: Looking back at Schmidt and Bjork (1992). *Perspectives on Psychological Science, 13*(2), 146–148. https://doi.org/10.1177/1745691617690642

Bui, G., Ahmadian, M. J., & Hunter, A-M. (2019). Spacing effects on repeated L2 task performance. *System, 81*, 1–13. https://doi.org/10.1016/j.system.2018.12.006

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin, 132*(3), 354–380. https://doi.org/10.1037/0033-2909.132.3.354

Challis, B. H. (1993). Spacing effects on cued-memory tests depend on level of processing. *Journal of Experimental Psychology, Learning, Memory, and Cognition, 19*(2), 389-396. https://doi.org/10.1037/0278-7393.19.2.389

Chang, A. C-S., & Millet, S. (2013). Improving reading rates and comprehension through timed repeated reading. *Reading in a Foreign Language, 25*(2), 126-148. https://doi.org/10125/66864

Cobb, T. (n.d.). Lextutor. Vocabprofile [computer program]. Accessed at https://www.lextutor.ca/vp/comp/

Conklin, K., Pellicer-Sánchez, A., & Carrol, G. (2018). *Eye-tracking: A guide for applied linguistics research.* Cambridge: Cambridge University Press.

DeKeyser, R. (2017). Knowledge and skill in ISLA. In S. Loewen & M. Sato (Eds.), *The Routledge handbook of instructed second language acquisition* (pp. 15–32). New York and London: Routledge.

Edmonds, A., Gerbier, E., Palasis, K., & Whyte, S. (2021). Understanding the

distributed practice effect and its relevance for the teaching and learning of L2 vocabulary. *Lexis: Journal in English Lexicology, 18*, 1-23. https://doi.org/10.4000/lexis.5652

Elgort, I. (2011). Deliberate learning and vocabulary acquisition in a second language. *Language Learning, 61*(2), 367–413. https://doi.org/10.1111/j.1467-9922.2010.00613.x

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*(2), 341–366. https://doi.org/10.1017/S0272263117000109

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

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. S*tudies in Second Language Acquisition, 35*(5), 483–517. https://doi.org/10.1017/S0272263113000119

Godfroid, A., Ahn, J., Choi, I., Ballard, L., Cui, Y., Johnston, S., et al. (2018). Incidental vocabulary learning in a natural reading context: An eye-tracking study. *Bilingualism: Language and Cognition, 21*(3), 563–84. https://doi.org/10.1017/S1366728917000219

Gorsuch, G., & Taguchi, E. (2008). Repeated reading for developing reading fluency and reading comprehension: The case of EFL learners in Vietnam. *System, 36*(2), 253–278. https://doi.org/10.1016/j.system.2007.09.009

Grabe, W. (2009). *Reading in a second language: Moving from theory to practice.* Cambridge: Cambridge University Press.

Han, Z., & Chen, C. A. (2010). Repeated-reading-based instructional strategy and vocabulary acquisition : A case study of a heritage speaker of Chinese. *Reading in a Foreign Language, 22*(2), 242–262. https://doi.org/10125/66838

Horst, M. & Meara, P. (1999). Test of a model for predicting second language lexical growth through reading. *The Canadian Modern Language Review, 56*(2), 308-328. https://doi.org/10.3138/cmlr.56.2.308

Hyönä, J., & Niemi , P. (1990). Eye movements in repeated movements of a text. *Acta Psychologica, 73*(3), 259 – 280. https://doi.org/10.1016/0001-6918(90)90026-C

IBM Corp. Released 2020. IBM SPSS Statistics for Windows, Version 27.0. Armonk, NY: IBM Corp.

Inhoff, A. W., Topolski, R., Vitu, F., & O'Regan, J. K. (1993) Attention demands during reading and the occurrence of brief (express) fixations. *Perception & Psychophysics, 54*(6), 814–823. https://doi.org/10.3758/BF03211806

Kim, Y. (2006). Effects of input elaboration on vocabulary acquisition through

reading by Korean learners of English as a foreign language. *TESOL Quarterly, 10*(2), 341–373. https://doi.org/10.2307/40264526

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*(1), 1103-1139. https://doi.org/10.1017/S0142716421000370

Koval, N. (2022). Testing the reminding account of the lag effect in L2 vocabulary learning. *Applied Psycholinguistics, 43*(1), 1-40. https://doi.org/10.1017/S0142716421000370

Küpper-Tetzel, C. E., Erdfelder, E., & Dickhäuser, O. (2014). The lag effect in secondary school classrooms: Enhancing students' memory for vocabulary. *Instructional Science, 42*(3), 373–388. https://doi.org/10.1007/s11251-013-9285-2

LaBerge, D., & Samuels, S. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 6*(2), 293–323. https://doi.org/10.1016/0010-0285(74)90015-2

Laufer, B. (2003) Vocabulary acquisition in a second language: Do learners really acquire most vocabulary by reading? Some empirical evidence. *Canadian Modern Language Review, 59*(4), 567–587. https://doi.org/10.3138/cmlr.59.4.567

Li, M., & DeKeyser, R. (2019). Distribution of practice effects in the acquisition and retention of L2 Mandarin tonal word production. *The Modern Language Journal, 103*(3), 607–628. https://doi.org/10.1111/modl.12580

Liu, Y. T., & Todd, A. G. (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

Llanes, À., & Tragant, E. (2021, online). Measuring the effects of repeated exposure to children's graded readers. *Language Teaching Research.* https://doi.org/10.1177/13621688211036609

Macmillan (n.d.). *Macmillan online dictionary.* Retrieved from https://www.macmillandictionary.com/

Meara, P. M., (1992). *EFL Vocabulary Tests.* Swansea: Centre for Applied Language Studies.

Meara, P. M. & Miralpeix, I. (2017). *Tools for researching vocabulary*. Clevedon: Multilingual Matters.

Nakata, T. (2015). Effects of expanding and equal spacing on second language vocabulary learning. *Studies in Second Language Acquisition, 37*(4), 677–711. https://doi.org/10.1017/S0272263114000825

Nakata, T. (2020). Learning words with flash cards and word cards. In S. Webb (Ed.), *The Routledge hanbook of vocabulary studies* (pp. 304-319). New York and London: Routledge.

Nakata, T., & Elgort, I. (2021). Effects of spacing on contextual vocabulary learning: Spacing facilitates the acquisition of explicit, but not tacit, vocabulary knowledge. *Second Language Research, 37*(2), 233-260. https://doi.org/10.1177/0267658320927764

Nakata, T., & Suzuki, Y. (2019a). Effects of massing and spacing on the learning of semantically related and unrelated words. *Studies in Second Language Acquisition, 41*(2), 287–311. https://doi.org/10.1017/s0272263118000219

Nakata, T., & Webb, S. (2016). Does studying vocabulary in smaller sets increase learning? The effects of part and whole learning on second language vocabulary acquisition. *Studies in Second Language Acquisition, 38*(3), 523-552. https://doi.org/10.1017/S0272263115000236

Nation, I. S. P. (1990). *Teaching and learning vocabulary.* Boston, MA: Heinle and Heinle.

Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher, 31*(7), 9-13.

Nation, I. S. P. (2013). *Learning vocabulary in another language.* Cambridge: Cambridge University Press.

Nation, I. S. P., & Webb, S. (2011). *Researching and analyzing vocabulary.* Boston, MA: Heinle and Heinle.

Nichols, W. D., Rupley, W. H., & Rasinski, T. (2008). Fluency in Learning to read for meaning: Going beyond repeated readings. *Literacy Research and Instruction, 48*(1), 1–13. https://doi.org/10.1080/19388070802161906

Pavlik, P. I., & Anderson, J. R. (2005). Practice and forgetting effects on vocabulary memory: An activation-based model of the spacing effect. *Cognitive Science, 29*(4), 559-586. https://doi.org/10.1207/s15516709cog0000_14

Pellicer-Sánchez, A. (2016). Incidental 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.

Pellicer-Sánchez, A., & Schmitt, N. (2012). Scoring *Yes-No* vocabulary tests: Reaction time vs. nonword approaches. *Language Testing, 29*(4), 489–509. https://doi.org/10.1177/0265532212438053

Peters, E., & Webb, S. (2018). Incidental vocabulary acquisition through viewing L2 television and factors that affect learning. S*tudies in Second language Acquisition, 40*(3), 551-577. https://doi.org/10.1017/S0272263117000407

Raney, G. E., & Rayner, K. (1995). Word frequency effects and eye movements

during two readings of a text. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 49*(2), 151–173. https://doi.org/10.1037/1196-1961.49.2.151

Rogers J., & Leow, R. (2020). Toward greater empirical feasibility of the theoretical framework for systematic and deliberate L2 practice: Comments on Suzuki, Nakata, & DeKeyser (2019). *Modern Language Journal, 104*(1), 309-312. https://doi.org/10.1111/modl.12621

Rohrer, D., & Pashler, H. (2007). Increasing retention without increasing study time. *Psychological Science, 16*(4), 183–186. https://doi.org/10.1111/j.1467-8721.2007.00500.x

Samuels, S. J. (2004). Toward a theory of automatic information processing in reading revisited. In R. B. Ruddell, M. R. Ruddell, & H. Singer (Eds.), *Theoretical models and processes of reading* (pp. 1127–1148). Newark, DE: International Reading Association.

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing, 18*(1), 55 - 88. https://doi.org/10.1177/026553220101800103

Schnitzer, B. S., & Kowler, E. (2006). Eye movements during multiple readings of the same text. *Vision Research, 46*(10), 1611-1632. https://doi.org/10.1016/j.visres.2005.09.023

Serrano, R. (2012). Is intensive learning effective? In C. Muñoz (Ed.), I*ntensive exposure experiences in second language learning* (pp. 3-22). Clevedon: Multilingual Matters.

Serrano, R., & Huang, H. (2018). Learning vocabulary through assisted repeated reading: How much time should there be between repetitions of the same text? *TESOL Quarterly, 52*(4), 971-994. https://doi.org/10.1002/tesq.445

Serrano, R., & Huang, H. (2023). Time distribution and intentional vocabulary learning through repeated reading: a partial replication and extension. *Language Awareness, 32*(1), 1-18. https://doi.org/10.1080/09658416.2021.1894162

Stephens, M. (2006). *New proficiency reading.* Essex: Pearson Education Limited.

Suzuki, Y. (2021). Optimizing fluency training for speaking skills transfer: Comparing the effects of blocked and interleaved task repetition. *Language Learning, 71*(2), 285-325. https://doi.org/10.1111/lang.12433

Suzuki, Y., & Hanzawa, K. (2022). Massed task repetition is a double-edged sword for fluency development. An EFL classroom study. S*tudies in Second Language Acquitision, 44*(2), 536-561. https://doi.org/10.1017/S0272263121000358

Suzuki, Y., & DeKeyser, R. (2017). Effects of distributed practice on the proceduralization of morphology. L*anguage Teaching Research, 21*(2), 166-188. https://doi.org/10.1177/1362168815617334

Suzuki, Y., Nakata, T., & DeKeyser, R. M. (2019). The desirable difficulty framework

as a theoretical foundation for optimizing and researching second language practice. *Modern Language Journal, 103*(3), 713-720. https://doi.org/10.1111/modl.12585

Suzuki, Y., Nakata, T., & DeKeyser, R. M. (2020). Empirical feasibility of the desirable difficulty framework: Toward more systematic research on L2 practice for broader pedagogical implications. *Modern Language Journal, 104*(1), 313-319. https://doi.org/10.1111/modl.12625

Therrien, W. (2004). Fluency and comprehension gains as a result of repeated reading: A metaanalysis. *Remedial and Special Education, 25*(4), 252–261. https://doi.org/10.1177/07419325040250040801

Thios S. J., & D'Agostino, P. R. (1976). Effects of repetition as a function of study-phase retrieval. *Journal of Verbal Learning and Verbal Behavior, 15*(5), 529-536. https://doi.org/10.1016/0022-5371(76)90047-5

Toppino T. C., & Bloom, L. C. (2002). The spacing effect, free recall, and two-process theory: A closer look. *Journal of Experimental Psychology: Learning, Memory and Cognition, 28*(3), 437-444. https://doi.org/10.1037/0278-7393.28.3.437

Toppino, T. C., & Gerbier, E. (2014). About practice: Repetition, spacing, and abstraction. In B. H. Ross (Ed.), *The psychology of learning and motivation: Volume 60: The psychology of learning and motivation* (pp. 113-189). San Diego, CA: Elsevier Academic Press.

UCLES (University of Cambridge Local Examination Syndicate) (2001). *Quick Placement Test.* Oxford: Oxford University Press.

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/10125/66776

Webb, S., & Chang, A. (2012). Vocabulary learning through assisted and unassisted repeated reading. *Canadian Modern Language Review, 68*(3), 267–290. https://doi.org/10.3138/cmlr.1204.1

Appendix A: Target words and distractors

Item	Type	Frequency band	Source
smock	target noun	11K	BNC-COCA 1-25K (Lextutor)
libel	target noun	7K	BNC-COCA 1-25K (Lextutor)
smock	target noun	11K	BNC-COCA 1-25K (Lextutor)
libel	target noun	7K	BNC-COCA 1-25K (Lextutor)
yam	target noun	11K	BNC-COCA 1-25K (Lextutor)
throng	target noun	7K	BNC-COCA 1-25K (Lextutor)
hype	target noun	6K	BNC-COCA 1-25K (Lextutor)
measles	target noun	10K	BNC-COCA 1-25K (Lextutor)
wacky	target adjective	10K	BNC-COCA 1-25K (Lextutor)
appalled	target adjective	4K	BNC-COCA 1-25K (Lextutor)
batty	target adjective	13K	BNC-COCA 1-25K (Lextutor)
shrewd	target adjective	6K	BNC-COCA 1-25K (Lextutor)
floppy	target adjective	5K	BNC-COCA 1-25K (Lextutor)
creased	target adjective	6K	BNC-COCA 1-25K (Lextutor)
twose	Pseudoword		Meara (1992)
scother	Pseudoword		Elgort (2011)
buttle	Pseudoword		Meara (1992)
wray	Pseudoword		Meara (1992)
preachet	Pseudoword		Elgort (2011)
obsolate	Pseudoword		Elgort (2011)
adair	Pseudoword		Meara (1992)
regrain	Pseudoword		Elgort (2011)
galpin	Pseudoword		Meara (1992)
mundy	Pseudoword		Meara (1992)
bance	Pseudoword		Meara (1992)
stace	Pseudoword		Meara (1992)
whim	low freq noun	9K	VST (Nation & Beglar, 2007)

rouble	low freq noun	13K	VST (Nation & Beglar, 2007)
skylark	low freq noun	13K	VST (Nation & Beglar, 2007)
blaspheme	low freq noun	10K	VLT (Nation, 1990)
beagle	low freq noun	13K	VST (Nation & Beglar, 2007)
peasantry	low freq noun	10K	VST (Nation & Beglar, 2007)
gauche	low freq adjective	14K	VST (Nation & Beglar, 2007)
canonical	low freq adjective	14K	VST (Nation & Beglar, 2007)
jovial	low freq adjective	13K	VST (Nation & Beglar, 2007)
limpid	low freq adjective	14K	VST (Nation & Beglar, 2007)
bawdy	low freq adjective	14K	VST (Nation & Beglar, 2007)
upbeat	low freq adjective	9K	VST (Nation & Beglar, 2007)
basket	high freq noun	2K	VLT (Nation, 1990)
standard	high freq noun	1K	VST (Nation & Beglar, 2007)
birth	high freq noun	2K	VLT (Nation, 1990)
patience	high freq noun	2K	VST (Nation & Beglar, 2007)
dinosaur	high freq noun	3K	VST (Nation & Beglar, 2007)
soldier	high freq noun	3K	VST (Nation & Beglar, 2007)
flesh	high freq noun	2K	VLT (Nation, 1990)
drawer	high freq noun	2K	VST (Nation & Beglar, 2007)
victory	high freq noun	2K	VLT (Nation, 1990)
stone	high freq noun	2K	VST (Nation & Beglar, 2007)
time	high freq noun	1K	VST (Nation & Beglar, 2007)
debt	high freq noun	2K	VLT (Nation, 1990)
holy	high freq adjective	2K	VLT (Schmitt, Schmitt & Clapham, 2001)
private	high freq adjective	2K	VLT (Nation, 1990)
total	high freq adjective	2K	VLT (Nation, 1990)
candid	high freq adjective	4K	VST (Nation & Beglar, 2007)
annual	high freq adjective	4K	VLT (Nation, 1990)
naked	high freq adjective	4K	VLT (Schmitt, Schmitt & Clapham, 2001)
upset	high freq adjective	2K	VST (Nation & Beglar, 2007)

ancient	high freq adjective	2K	VLT (Schmitt, Schmitt & Clapham, 2001)
aware	high freq adjective	3K	VLT (Schmitt, Schmitt & Clapham, 2001)
original	high freq adjective	2K	VLT (Nation, 1990)
lovely	high freq adjective	2K	VLT (Schmitt, Schmitt & Clapham, 2001)
poor	high freq adjective	1K	VST (Nation & Beglar, 2007)

Appendix B: Vocabulary pretest

Please mark "YES" if **you know the meaning** of the following words, and mark "NO" if you don't know the meaning (no guessing). Keep in mind that there are non-words (for which you'll have to mark "NO", obviously), so please be honest and mark "NO" when applicable.

	YES	NO
twose		
holy		
smock		
wacky		
private		
total		
basket		
scother		
appalled		
buttle		
wray		
batty		
preachet		
shrewd		
whim		
candid		
standard		
floppy		
rouble		
skylark		
creased		
libel		
birth		
yam		
blaspheme		
annual		
throng		
canonical		
patience		
dinosaur		

	YES	NO
beagle		
naked		
obsolate		
peasantry		
soldier		
upset		
ancient		
flesh		
hype		
aware		
adair		
drawer		
jovial		
victory		
limpid		
stone		
original		
lovely		
regrain		
galpin		
gauche		
time		
poor		
mundy		
bance		
stace		
measles		
bawdy		
upbeat		
debt		

Appendix B: Vocabulary immediate and delayed posttest

Multiple choice test. Please choose the meaning of the following words:

1. smock
   a) a long loose shirt
   b) false appearance
   c) a type of facial cream
   d) determination
   e) I don't know
2. libel
   a) a corrupt plan, especially for getting money
   b) the illegal act of writing things about someone that are not true
   c) holding an opinion and sharing it without careful thought
   d) the act of stealing money that people trust you to look after as part of your work
   e) I don't know
3. hype
   a) something you say that is not true
   b) the act of emphasizing what you're saying
   c) sounds made by voices or instruments
   d) the use of a lot of publicity to influence people
   e) I don't know
4. throng
   a) a large number of people crowded or assembled together
   b) someone whose job is to report the news for a newspaper
   c) a small group of people who have a lot of advantages
   d) someone who belongs to an organization
   e) I don't know
5. yam
   a) a sweet sticky food made from boiled fruit and sugar
   b) an African tree known for its sweet, yellow fruit
   c) a root vegetable that looks like a long white potato
   d) large grey bird, originally from Africa
   e) I don't know
6. measles
   a) an extended shortage of a basic nutrient
   b) large groups of insects flying or moving together
   c) an uncontrolled increase in the numbers of an insect
   d) an infectious disease characterized by red spots all over the body
   e) I don't know

7. creased
    a) full of lines
    b) not pale in colour
    c) completely even
    d) full of brown spots
    e) I don't know
8. appalled
    a) obvious or very easily noticed
    b) hurt by hitting or kicking
    c) offended or shocked
    d) with skin that is lighter than usual
    e) I don't know
9. floppy
    a) done in a very careless way
    b) spending or costing a lot of money
    c) lacking power or influence
    d) soft and hanging down in a loose way
    e) I don't know
10. batty
    a) careful and using good judgment
    b) not able to see clearly
    c) slightly crazy
    d) refusing to listen to other opinions
    e) I don't know
11. shrewd
    a) clever and able to make good judgments
    b) behaving in a way that is not polite
    c) morally bad, dangerous, or frightening
    d) caring about other people more than about oneself
    e) I don't know
12. wacky
    a) famous or successful
    b) morally wrong
    c) funny or silly
    d) not strong or definite
    e) I don't know