


ARTICLE

Testing the deficient processing account of the spacing effect in second language vocabulary learning: Evidence from eye tracking

Natalie G. Koval* 

Michigan State University

*Corresponding author. E-mail: kovalnat@msu.edu

(Received 14 May 2018; revised 12 February 2019; accepted 19 February 2019)

Abstract

The spacing effect refers to the learning benefit that comes from separating repeated study of target items by time or by other items. A prominent proposed explanation for this effect states that repeated exposures that occur closely together may not engage full attentional processing due to residual activation of the previous exposure and also, in an intentional learning context, due to a sense of familiarity that may result in strategic allocation of less study time to an item in massed repetitions. The present study used eye-tracking methodology to investigate the effects of temporal distribution of repeated exposures to novel second language words on attentional processing and learning of these words under intentional learning instructions. Adult native speakers of English read Finnish words embedded in English sentence contexts under massed and spaced conditions. The results showed that (a) massed repeated exposures received less attentional processing than spaced repeated exposures; (b) target words were better remembered in the spaced condition; and (c) attention was a significant mediator of the obtained spacing effect, in line with the predictions of the deficient processing account of the spacing effect. Implications for vocabulary learning are discussed.

Keywords: attention; eye tracking; second language vocabulary acquisition; spacing effect; study-phase retrieval

It is widely believed that attention plays an important role in second language (L2) learning (Gass, 1988; Robinson, 2003; Schmidt, 1990, 2001). This includes learning L2 vocabulary. Studies employing eye-tracking methodology have consistently found a positive relationship between the time a reader spends looking at a novel word and learning of the word (Godfroid et al., 2017; Godfroid, Boers, & Housen, 2013; Godfroid & Schmidtke, 2013; Mohamed, 2017; Pellicer-Sánchez, 2016). Such studies further show that amount of attentional processing benefits learning above and beyond the number of exposures to a target word (Godfroid et al., 2017; Mohamed, 2017), the latter variable being one of the best known and most intuitive

predictors of vocabulary learning success (Webb, 2007). Thus, we know that vocabulary learning is more successful if the to-be-learned words are encountered repeatedly and also if these encounters are processed more attentively.

A number of studies have used eye tracking to investigate how repeated exposures to novel vocabulary are processed in terms of the amount of attention they receive and how this affects learning. A variable that has not yet been considered in this line of inquiry is the temporal distribution of repeated exposures. However, this variable is important because how closely together or widely apart repeated encounters with a target word occur may have an effect on both attention and learning. Research in psychology has consistently shown that learning from repeated study is more effective when repetitions are distributed over time rather than massed together (Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006). This is known as the spacing effect. A prominent explanation for the effect, known as the deficient processing theory, states that repeated encounters with to-be-learned material receive more attentional processing if they are spaced more widely (Callan & Schweighofer, 2010; Cuddy & Jacoby, 1982). Because attention is known to be important for L2 learning success, it is important to understand how it is affected by the temporal distribution of repeated exposures to target forms. The present study extends on previous research into attentional processing and learning during repeated exposures to novel L2 vocabulary by including a manipulation of the temporal distribution of exposures. If, as predicted by the deficient processing hypothesis, the more widely spaced exposures are found to engage more attentional processing, this may have important implications for efforts to induce more attention to target forms in L2 pedagogy. Further, with the help of a mediation analysis, I investigate the contribution of deficient processing to the spacing effect in L2 vocabulary learning from different sentence contexts.

The Spacing Effect

The spacing effect (or *distributed practice effect*) refers to the widely obtained finding in psychology that memory for repeatedly studied material is better when repetitions are separated by time or intervening material than when the same number of repetitions follow consecutively, or in a massed fashion (Cepeda et al., 2006; Cepeda, Vul, Rohrer, Wixted, & Pashler, 2008; Delaney, Verhoeven, & Spigler, 2010; Dempster, 1988; Gerbier & Toppino, 2015; Pavlik & Anderson, 2005; Rohrer & Pashler, 2007). A related finding termed *the lag effect* is the finding that longer interstudy intervals are more beneficial for long-term retention of knowledge than shorter interstudy intervals (Melton, 1967). Probably because the two effects are so closely related and also because in some situations “massed” practice in its strictest sense may not be as relevant as practice that is separated by intervals of varying lengths (Seabrook, Brown, & Solity, 2005), the terminological distinction is not always made in the literature, where the term *spacing effect* is often used to refer to lag effects (see, e.g., Bird, 2010; Nakata & Webb, 2016; Zhao et al., 2015). For simplicity, the term *spacing effect* will be used henceforth to refer to both phenomena.

The spacing effect is one of the most robust findings in memory research (Cepeda et al., 2006; Gerbier & Toppino, 2015), where it has been the focus of much interest since at least the late 1800s (Ebbinghaus, 1885). The benefits of spacing practice have been consistently obtained in a wide range of experimental paradigms, with vastly diverse tasks, and with vastly diverse populations that are not limited to our species (Donovan & Radosevich, 1999; Yin, Del Vecchio, Zhou, & Tully, 1995). The spacing effect has also consistently been obtained with foreign language vocabulary as learning targets (see, e.g., Bahrick, Bahrick, Bahrick, & Bahrick, 1993; Callan & Schweighofer, 2010; Kang, Lindsey, Mozer, & Pashler, 2014; Pashler, Zarow, & Triplett, 2003; Pavlik & Anderson, 2005). The robustness and generality of the effect suggests that it potentially holds great promise for any learning situation. In order to make the best use of this powerful learning tool, we need a good understanding of the variables that may mediate and moderate the relationship between the spacing of repeated study and learning outcomes in our specific learning contexts. However, second language acquisition (SLA) research has thus far focused only on whether or not spacing is beneficial for the acquisition of various aspects of an L2, such as syntax (Bird, 2010; Rogers, 2015) and vocabulary (Bloom & Shuell, 1981; Nakata, 2015; Nakata & Webb, 2016; Schuetze, 2015). No SLA studies, to my knowledge, have yet investigated the associated underlying mechanisms. The present study begins to address this gap by exploring the contribution of deficient processing as a potential mediator for the spacing effect in contextual learning of novel L2 vocabulary. It is recognized today that likely more than one cognitive mechanism underlies the spacing effect and that the engagement of these different processes or combinations of processes may depend on a number of variables, such as the target task (Gerbier & Toppino, 2015; Glenberg & Smith, 1981; Greene, 1989; Kornell & Bjork, 2008; Russo & Mammarella, 2002). A good understanding of what processes are relevant for SLA contexts is crucial for determining how and when the spacing effect may be useful for L2 teaching and learning as well as how to use it most effectively in any particular language learning situation (Rogers, 2015, p. 864). The focus of this study on deficient processing as a potential mediator for the spacing effect in L2 vocabulary acquisition is motivated by the widely accepted theory that attention facilitates L2 learning (Schmidt, 1990, 2001). The next section briefly discusses the deficient processing theory as well as other prominent efforts in psychology research to specify the mechanisms underlying the spacing effect.

Explaining The Spacing effect

A large number of theories have been proposed to explain the spacing effect (see, e.g., Benjamin & Tullis, 2010; Bjork & Allen, 1970; Challis, 1993; Estes, 1955; Glenberg, 1979; Hintzman, Block, & Summers, 1973; Jacoby, 1978; Landauer, 1969; Maddox, 2016; Melton, 1970; Pavlik & Anderson, 2005; Raaijmakers, 2003). Theories that have received the most empirical attention can be broadly classified into those that explain the effect in terms of encoding variability and those that explain it in terms of deficient processing. According to the encoding variability family of accounts (Bower, 1972; Glenberg, 1976; Greene, 1989; Maddox, 2016; Raaijmakers, 2003), spaced repetitions are more likely to be experienced in different

contextual states, resulting in the encoding of more diverse contextual elements that serve as important retrieval routes at test. According to the deficient processing family of accounts (Callan & Schweighofer, 2010; Challis, 1993; Cuddy & Jacoby, 1982; Dellarosa & Bourne, 1985; Gerbier & Toppino, 2015; Greene, 1989; Hintzman et al., 1973; Jacoby, 1978; Johnston & Uhl, 1976; Krug, Davis, & Glover, 1990; Pavlik & Anderson, 2005; Zechmeister & Shaughnessy, 1980), when an item is repeated immediately or very shortly after its initial presentation, the repetition is processed less fully because its prior presentation may still be activated in short-term memory or readily accessible (Cuddy & Jacoby, 1982; Greeno, 1967; Whitten & Bjork, 1977). The difference in the amount of processing that an item receives is said to be responsible for the observed differences in learning outcomes.

Deficient processing of massed repetitions may be due to voluntary mechanisms, where it is the result of a conscious decision to devote less attention or study time to an immediately repeated item due to an increased sense of familiarity (Greene, 1989; Rundus, 1971; Shaughnessy, Zimmerman, & Underwood, 1972; Verkoijen & Delaney, 2008; Zechmeister & Shaughnessy, 1980; Zimmerman, 1975). Alternatively, a failure to process a stimulus as extensively when it is repeated immediately may be due to more automatic mechanisms of a less strategic nature (Callan & Schweighofer, 2010; Challis, 1993; Mammarella, Avons, & Russo, 2004; Topino, 1991; Van Strien, Verkoijen, Van der Meer, & Franken, 2007; Xue et al., 2011). Processing may further be deficient in terms of quantity or quality. In the former case, the same underlying operations are engaged to a lesser extent or simply operate more quickly. In the latter case, the system relies on a different set of operations or some of the encoding operations are dropped and are not repeated during the second processing event. An example of a qualitative change in processing is retrieving a solution to a problem from memory instead of repeating the computational process that produced the solution (Jacoby, 1978; Rose, 1984).

The generality of the spacing effect suggests that it likely reflects some fundamental property of the memory system (Gerbier & Toppino, 2015, p. 50). However, specifying a single unitary mechanism that would account for the wide range of consistently obtained findings has proven difficult (Benjamin & Tullis, 2010; Delaney et al., 2010; Gerbier & Toppino, 2015; Greene, 1989; Maddox, 2016; Verkoijen, Rikers, & Schmidt, 2004). Thus, in their basic form, the encoding variability and deficient processing theories cannot readily accommodate the important findings of super-additivity and non-monotonicity (Benjamin & Tullis, 2010; Maddox, 2016). *Super-additivity* refers to the finding that likelihood of successful performance on a memory test following spaced practice is greater than what would be expected from independent encoding events (Begg & Green, 1988; Benjamin & Tullis, 2010; Ross & Landauer, 1978). This suggests that repetition plays a special role and a dependency between the memory traces is crucial (Bellezza, Winkler, & Andrasik, 1975; Hintzman & Block, 1973). Such trace dependency is not readily accommodated by the encoding variability or deficient processing theories in their basic form, as the more independent repetitions are, the more likely they are to be encoded with different contextual associations and to be processed fully.

Non-monotonicity refers to the shape of the lag function (an inverted U). Thus, while increasing the interstudy interval initially has a positive effect on learning, there comes a point at which an optimal interstudy interval is reached and beyond

which increasing spacing any further actually has a negative effect on memory (Cepeda et al., 2006, 2008). As both contextual variability and amount of processing should monotonically increase the longer the interstudy interval, this finding, too, cannot be readily explained.

To better account for experimental evidence, a number of multiprocess theories have been proposed. These (see, e.g., Benjamin & Tullis, 2010; Delaney et al., 2010; Greene, 1989; Maddox, 2016; Verkoijen et al., 2004) combine the effects of contextual fluctuation or deficient processing with the central assumption of the study-phase retrieval theory (Delaney et al., 2010; Greene, 1989; Hintzman et al., 1973; Madigan, 1969; Thios & D'Agostino, 1976). The assumption is that an item's initial presentation must be retrieved at the time of its repetition, or that the repetition of a stimulus must "remind" (Benjamin & Tullis, 2010) of its previous occurrence. Both super-additivity and non-monotonicity can be accommodated with this additional assumption. At very long interstudy intervals, successful retrieval is less likely due to memory trace decay, which would lead to a failure of super-additive effects and, consequently, less effective learning. This, in turn, would show up as the postinflection downward trajectory in learning as a function of spacing.

Other important findings can also be accommodated if we make this additional assumption. An important finding is that intentional learning produces larger spacing effects and also shows a longer optimal interstudy interval than incidental learning (Toppino & Bloom, 2002; Verkoijen, Rikers, & Schmidt, 2005). This makes sense if we assume that intentional learning produces stronger and more durable memory traces. Another important finding is that when the context of study is purposefully changed between repetitions, such experimenter-introduced variability benefits learning of massed repetitions but has an adverse effect on learning of spaced repetitions (Johnston & Uhl, 1976; Verkoijen et al., 2004). If we assume that a repetition must remind of its previous presentation, then the detrimental effects of experimenter-imposed context change for the spaced condition can be explained by the fact that such change may render repetitions too dissimilar and the increased lag may further make recognition less likely (Benjamin & Tullis, 2010; Verkoijen et al., 2004).

Investigating Deficient Processing

Repetitions that occur close in time have been consistently shown to engage less attentional processing relative to spaced repetitions in a number of experimental paradigms in psychological research. It has been shown, for example, that less time is devoted to processing or studying stimuli and performing tasks that are repeated immediately, including repeated readings of texts (Hyönä & Niemi, 1990; Krug et al., 1990; Rose, 1984; Wahlheim, Dunlosky, & Jacoby, 2011). Massed repetitions of task performance require less effort as measured by pupil dilations (Magliero, 1983) and secondary task performance (Johnston & Uhl, 1976). Further evidence that massed repetitions receive diminished processing comes from studies that have used techniques such as electroencephalogram and functional magnetic resonance imaging to investigate brain activity during the processing of massed and spaced repetitions (see, e.g., Callan & Schweighofer, 2010; Kim, Kim, & Kwon, 2001; Van Strien et al., 2007; Xue et al., 2011; Zhao et al., 2015).

Past research into deficient processing as an explanation for the spacing effect has also employed self-paced study of stimuli, such as words (Rundus, 1971; Shaughnessy et al., 1972; Zimmerman, 1975). Here, target words are presented one per slide, and participants move from slide to slide at their own pace. The time participants spend on each slide is recorded as a measure of processing time. These studies showed that participants allocated more study time to more widely spaced words and these, in turn, were better remembered. However, because the benefit of spacing was greater than what was predicted from study time, Shaughnessy et al. speculated that participants may have used a strategy of holding off on ending a trial not to appear uninterested, particularly with massed repetitions. Some studies have included recordings of overt rehearsal (Rundus, 1971, Experiment 3; Zimmerman, 1975). Zimmerman (1975) included this specifically to ensure equivalence between nominal and functional study time, reasoning that as long as participants are saying a word aloud, some type of processing must still be going on. However, learning benefits were still underestimated by study time. This led Zimmerman to question the ability of deficient processing to serve as an explanatory mechanism for the spacing effect. It could be argued, however, that overt rehearsal may still not capture certain important differences in processing, such as the fact that the nature of processing may not be constant throughout, with the first time a word is read aloud (either at initial exposure or after a delay) engaging deeper processing than when it is repeated toward the end of a series of consecutive repetitions. Further, this design, too, is not immune to effects of strategies as, in order to avoid appearing uninterested, participants may still repeat a word beyond what they might normally do, which might, again, be more likely with massed repetitions.

The present experiment extends on this line of research by using eye tracking to investigate self-paced study of novel L2 words within sentence contexts. Eye-tracking methodology consists in recording what parts of written input a participant looks at and for how long as well as the progression of their eye movements through written discourse. Eye tracking affords millisecond precision in measuring attention allocation to different parts of the visual display. Here, exposure time and processing time are more likely to be equivalent because of the hypothesized tight “eye-mind link” (Just & Carpenter, 1980). Eye-tracking methodology further affords an in-depth investigation into the different stages of processing, allowing us to look separately at early processes involved in word recognition and later processes that may include intentional rehearsal. The many different eye-movement indices that are recorded can further inform about qualitative differences in attention patterns between experimental conditions, which also allows us to infer any strategies adopted by participants in performing a given task.

While the overwhelming majority of spacing effect studies have investigated rote memorization, such as learning of lists of paired associates, benefits of spacing have also been demonstrated with more meaningful materials and higher level learning (Helsdingen, van Gog, & van Merriënboer, 2011; Kapler, Weston, & Wiseheart, 2015; Kornell & Bjork, 2008; Reder & Anderson, 1982; Sobel, Cepeda, & Kapler, 2011; Vlach & Sandhofer, 2012; Wahlheim et al., 2011). It has also been shown that the relevance of deficient processing theories is not limited to rote learning (Krug et al., 1990; Rose, 1984; Wahlheim et al., 2011). In the present study, participants learned novel L2 words within sentence contexts, which constitutes more meaningful

learning and is a departure from the traditional word list learning paradigm. Further, repetitions of the same word occurred in different sentence contexts. Encountering target vocabulary in different contexts is a common scenario for language learning. Such context variability has been shown to have the effect of increasing processing for massed repetitions (Dellarosa & Bourne, 1985). Thus, in the present study massed repetitions cannot be processed as inattentively as they might in word list learning because each new sentence context necessitates processing of the target word in a new situation for comprehension of the sentence. However, for spaced repetitions such variability might mean lower likelihood of study-phase retrieval, or reminding (Benjamin & Tullis, 2010; Verkoijen et al., 2004), which might have a negative effect on learning. The present study investigates whether spacing will still have the effect of increasing attention above and beyond what is engendered by such variety in contexts and whether benefits of spacing will be observed in such diverse sentence contexts.

Attention and L2 Vocabulary Learning

In both incidental and intentional L2 vocabulary learning, more attentional engagement with a word has been shown to result in better learning outcomes (Craik & Lockhart, 1972; Godfroid et al., 2013; Laufer & Hulstijn, 2001; Mohamed, 2017; Schmitt, 2008). Studies employing eye-tracking methodology have consistently shown a positive relationship between the amount of attention a target word receives and learning of the word (Godfroid et al., 2013; Godfroid et al., 2017; Godfroid & Schmidtke, 2013; Mohamed, 2017; Pellicer-Sánchez, 2016). From among the different measures that are used in eye tracking, attentional processing of vocabulary items during reading is investigated by measuring the duration and number of fixations that an area of interest receives. A fixation is when the gaze remains relatively still on an area of interest. It is during fixations that visual information is taken in (Rayner, 2009). Fixation duration is indicative of amount of attentional processing and encoding effort (Rayner, 1998). An area of interest often receives more than one fixation: these may be consecutive or they may be the result of a regression, which refers to revisiting an area of interest after the gaze has moved past it. In investigating novel word processing, the following measures are often employed: first fixation duration, which is the duration of the very first fixation on an area of interest; gaze duration, which is the sum of all fixations in an area of interest before the eyes exit the area of interest either to the right or to the left; and total reading time, which is the sum of all fixations in an area of interest, including those made during regressions.

A number of studies have used eye tracking to explore the effects of repetition on attention and learning of novel words (Elgort, Brysbaert, Stevens, & Van Assche, 2017; Godfroid et al., 2017; Joseph, Wonnacott, Forbes, & Nation, 2014; Mohamed, 2017; Pellicer-Sánchez, 2016). In addition to reporting a positive relationship between attention and learning, these have consistently reported a decrease in reading times across repetitions, which is taken as an indication of increased familiarity and, in some studies, even used as a measure of learning (e.g., Joseph et al., 2014). Another very important finding from this line of research is that amount of attention benefits learning above and beyond the number of times a target word is

encountered (Godfroid et al., 2017; Mohamed, 2017). This suggests that learning success depends not only on how many times a learner is exposed to a word but also on how these repeated exposures are utilized, or how much attentional processing they receive. Thus, finding ways of inducing more attention to target words is an important endeavor that potentially holds great benefits for teaching and learning vocabulary. Existing research offers valuable insights into how repeated encounters with novel vocabulary are processed and how this, in turn, affects learning. Missing from this line of research, however, is a consideration of any effects that temporal distribution of repeated encounters may have on both learning and attention, as well as, possibly, on the relationship between the two. The present study extends on existing research by investigating the effects of this important variable.

The Present Study

The present study used a within-subjects within-items counterbalanced design to explore the following research questions:

1. Does repetition affect attentional processing of novel L2 words differently under massed and spaced conditions?
2. Does the massing versus spacing of repeated exposures to novel L2 vocabulary affect learning gains?
3. Is the effect of spacing, if any, mediated by attention, operationalized as reading times for the target words?

Participants

Participants were 40 undergraduate and graduate students (21 males and 19 females, age: 18–42 years, $M = 21$, $SD = 4.4$) from a wide range of majors at Michigan State University, selected randomly from a large number of students who responded to an ad for the study that had been placed through the office of the registrar. All had normal or corrected-to-normal vision. All were native speakers of English. Ninety percent indicated that they had knowledge of a language other than English, with Spanish being reported most often (59%). Thirty-two percent indicated knowledge of more than one foreign language. None had any knowledge of Finnish (the language used in the present study). All were compensated for their time with cash. The sample size was based on Brysbaert and Stevens's (2018) recommendation for properly powered repeated-measures reaction time experiments that use mixed-effects statistical models.

Materials

The experiment consisted of a study phase, a distractor math task, immediate vocabulary posttests, delayed vocabulary posttests, and a background questionnaire.

Study materials

Twenty-four Finnish words were selected as targets. These were chosen such that they were not cognates of their English translations. The words were divided into two lists, one of which was to be presented in the massed and the other in the spaced

conditions, in a counterbalanced design. List A contained 9 nouns and 3 adjectives. List B contained 11 nouns and 1 adjective. The two lists were matched on the number of letters per word: each had three five-letter words, four six-letter words, one seven-letter word, three eight-letter words, and one nine-letter word. All target words denoted simple concepts (Appendix A), whose English translations were within the 5,000 most frequent words according to the BYU Corpus of Contemporary American English (Davies, 2008–).

The Finnish words were embedded in English sentence contexts. This mimicked what is known as the Clockwork methodology (Godfroid et al., 2017; Horst, Cobb, & Meara, 1998). This allowed me to target simple word learning characteristic of novice L2 vocabulary learning stages while using contexts that are more interesting to read. It further helped to control for any effects of sentence reading disfluency spilling over on target words and thus allowed me to better isolate the effects of interest. The target words were never the first or last word of a sentence. The sentences were simple in structure and meaning. Four different sentence contexts were created for each target word for a total of 96 experimental sentences. In addition, there were 24 buffer sentences that did not contain a Finnish word.

Twenty-five percent of the sentences were followed by comprehension questions, many of which followed buffer sentences. Only 10 of the experimental sentences (5 per list) were followed by a comprehension question. In the massed condition, the questions always came after the fourth repetition to preserve the massed nature of presentation. In the spaced condition, comprehension questions occurred across repetitions such that they, as was the case in the massed condition, were dispersed more or less evenly throughout the study phase. None of the questions contained a target word translation to avoid causing additional processing of the target word or its meaning. The purpose of the comprehension questions was to ensure that participants read the sentences for meaning. A participant had to answer at least 80% of the questions correctly for his or her data to be included.

Distractor math task

The math task was a sequence of self-paced paper-and-pencil mathematical equations performed during breaks in the experimental procedure (the 6 min between the study blocks and the 15 min between the study phase and the immediate posttests). A math task was used to fill these time intervals so that the type of cognitive activity in which the participants engaged would be roughly similar across participants. They were asked to prioritize accuracy of answers over speed. They were not allowed to use a calculator but were provided with additional sheets to use as scratch paper.

Vocabulary posttests

Two paper-based tests, a form recognition test and a form–meaning mapping test, were used to measure learning gains. The form recognition test measured learners' ability to recognize the correct spelling of each target word from among four choices. The three distractor items were created by switching the order of the first two syllables and letters within the syllables to produce plausible Finnish words from a nonnative speaker perspective. Whenever such a scrambling pattern resulted

in a more salient distractor, such as when a letter was doubled or the resulting non-word resembled an English word, a different scrambling technique was used, such as using the third syllable for scrambling. The test was presented printed on a single sheet of paper (test sheet 1), where each entry was numbered and included the target word, three distractors, and an “I don’t know” option. The same form recognition test (but with a different randomization) served as the immediate and delayed tests.

The form–meaning mapping test measured learners’ ability to match the target forms with their meanings. With a pen of a different color (to prevent changes to test sheet 1 at this point), the number for each entry on test sheet 1 was to be written next to its corresponding English translation on test sheet 2. Test sheet 2 contained 45 English words. Only 24 of these were the target English words; distractor words were taken from the buffer and practice sentences; thus, all English words on test sheet 2 had been encountered during the study phase. The reason test sheet 1 was used instead of a new sheet with only correct forms on it was to prevent additional exposure to the correct forms, which might have contaminated the results of the delayed form recognition test. Because temporal distribution of practice was of primary interest, I wanted to avoid any spaced practice for the items in the massed condition. Further, because participants would have just finished selecting target forms from among distractors, presenting them with a list of correct forms might serve as a learning opportunity at a time of heightened curiosity about the correct forms. This was confirmed in the piloting phase. The immediate and delayed form–meaning mapping tests were also identical except for order randomization. One point was awarded for each correct and zero points for each incorrect answer on both tests for a total of 24 possible points on each test.

Background questionnaire

A paper-based background questionnaire asked the participants to list all languages they had ever studied or had any knowledge of and to state whether or not any of the studied words had struck them as familiar upon initial exposure.

Procedure

The experiment was conducted in an eye-tracking lab, where each participant met individually with the researcher for two sessions with 48–72 hr in between. The study phase and the immediate posttests were completed in Session 1; the delayed posttests were completed in Session 2. First, each participant read the consent form and indicated his or her consent to participate. Then participants read the on-screen instructions and did a practice block consisting of eight trials. After the practice block, the researcher confirmed with the participant that all was clear before beginning the four experimental blocks of the study phase. The four blocks were separated by three 6-min breaks during which the participants performed the distractor math task.

The Finnish words in the massed condition were repeated in 4 consecutive sentences within the same block. There were three such words per block. All words in the spaced condition occurred in 1 sentence in each block. In order to keep spacing of the repetitions more or less constant in the spaced condition, the order of their

<p>BLOCK 1</p> <p>The plane was supposed to leave at nine but there was a delay.</p> <p>It is obvious that this rakennus is very old.</p> <p>I walk past this rakennus on my way to work every morning.</p> <p>Construction of the rakennus was paused until the issue was resolved.</p> <p>His apartment is in a tall rakennus on a downtown street.</p> <p>There is no silta across this river; you have to take the ferryboat.</p> <p>They gave us a tour around the kaupunki and took us shopping.</p> <p>He sat on a bench smoking a savuke and reading a newspaper.</p> <p>This is a tarina written by my brother.</p> <p>I don't think this osoite is correct.</p> <p>...</p> <p>6 minutes of distractor math task</p>	<p>BLOCK 3</p> <p>Dan hasn't eaten anything since this morning.</p> <p>I don't think this paita goes with those pants.</p> <p>I don't know why my paita is hanging in his closet.</p> <p>Fold the paita carefully so it doesn't wrinkle.</p> <p>Michelle gave Bob a paita for Christmas.</p> <p>Andrew offered me a savuke but I don't smoke.</p> <p>I want to hear that tarina again, please.</p> <p>The workers are building a silta that will connect the two points.</p> <p>One of the founders of the kaupunki was a famous physician.</p> <p>He saw paw prints on the lattia that led to the kitchen.</p> <p>...</p> <p>6 minutes of distractor math task</p>
<p>BLOCK 2</p> <p>Richard and Emma are getting a divorce.</p> <p>There are many stores in our kaupunki that sell this kind of stuff.</p> <p>Driving across the silta he admired the beauty of the nature around.</p> <p>The last tarina in this book is very interesting.</p> <p>A stranger asked if he had a savuke and he offered him one.</p> <p>The wings of this perhonen have beautiful patterns on them.</p> <p>Not all species of perhonen like sunlight.</p> <p>He spotted a colorful perhonen sitting on a leaf.</p> <p>He opened the jar but the perhonen did not fly out.</p> <p>The grammar of this kieli is very complex.</p> <p>...</p> <p>6 minutes of distractor math task</p>	<p>BLOCK 4</p> <p>Robert moved to New York after graduating from college.</p> <p>Frank reads his kids a tarina every night.</p> <p>He pulled a savuke out of his pocket and lit it.</p> <p>He lives in another kaupunki and commutes to work every day.</p> <p>We were told that the silta was closed for repair.</p> <p>I watched the elokuvu that you suggested.</p> <p>I think that this elokuvu is much scarier than that one.</p> <p>They made a funny elokuvu about his adventures.</p> <p>We will watch a short elokuvu at the beginning of class.</p> <p>The article was written in a kieli that he didn't know.</p> <p>...</p> <p>15 minutes of distractor math task</p>

Figure 1. An illustration of the target word distribution in the study phase. The figure presents 10 sentences at the beginning of each block. Examples of the occurrence distribution for one word in the massed condition (repeating in 4 consecutive sentences) and one word in the spaced condition (repeating in 4 sentences across the four blocks) are highlighted.

occurrence was changed only slightly from block to block, which was done to prevent any (however unlikely) anticipation of their occurrence across the four blocks. Each block started and ended with 1 buffer sentence. The final block always ended with an additional 16 buffer sentences. These served to prevent words in target sentences close to the end of the study phase from having a memorial advantage. Thus, words in the spaced condition were repeated in sentences separated by 25 other sentences and the distractor task. Figure 1 illustrates the distribution of the target words in the study phase. The carrier sentences are highlighted for two of the words, one in the massed condition (it repeats in 4 sentences that are consecutive) and one in the spaced condition (it repeats in 4 sentences across the four blocks). Thus, words in the massed and spaced conditions occurred throughout the four blocks of the experiment. Their average serial position was controlled ($M = 60.50$ and 60.54 for massed and spaced words, respectively, $p = .972$).

Each block took 9–17 min to complete ($M = 12.22$, $SD = 1.73$), which together with the 6 min of the distractor task made for a 15- to 23-min interstudy interval for the spaced condition ($M = 18.22$, $SD = 1.73$). Experimental blocks were rotated such that all blocks appeared in each of the four positions, which, together with the counterbalancing between conditions, resulted in eight different experimental sequences, one sequence per five participants.

Participants read the study phase sentences from a computer screen while their eye movements were recorded with the EyeLink 1000 eye tracker (SR Research Ltd.; Mississauga, Canada), which samples at 1000 Hz. While reading, participants were seated in front of the computer monitor at a distance of 66 cm from the screen, where one letter subtended 0.36 degree of horizontal visual angle. To minimize head movements, participants rested their chin and forehead against a head stabilizer.

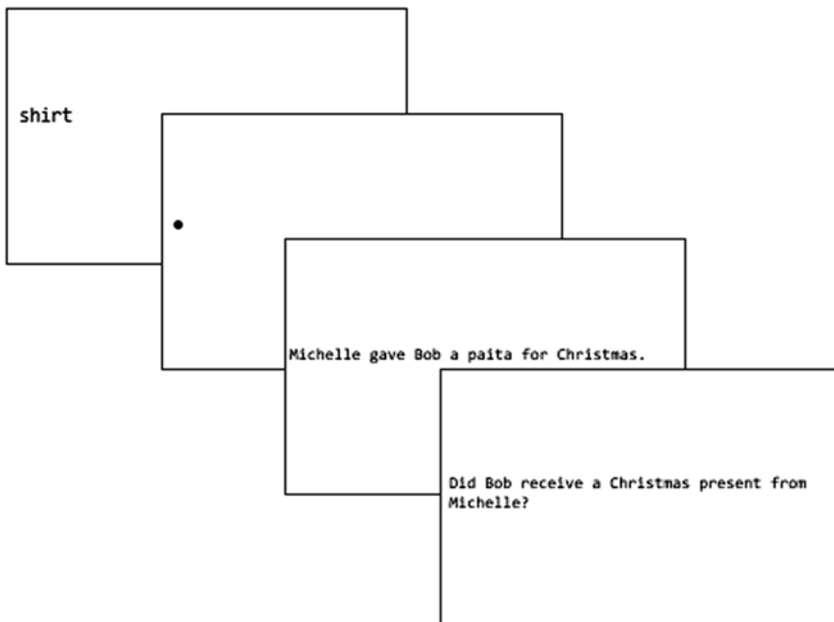


Figure 2. Sample study-phase trial sequence with a comprehension question. The English translation for the upcoming Finnish word is always presented before the target sentence.

The sentences were presented in black Consolas font, size 18, against a light-gray background. While reading was binocular, only the movements of the right eye were monitored. Each sentence fit on a single line. A 9-point grid calibration procedure was performed before each block of the experiment and repeated as needed, such that the resulting calibration error was always less than 0.5 degree of visual angle.

The instructions informed the participants that there would be a test on their memory for the form and meaning of the new Finnish words at the end, as in a usual foreign language classroom. They were instructed not only to read each sentence for comprehension but also to try to learn the Finnish words as if this was a Finnish lesson in which words were learned in sentence contexts. Each experimental trial started with the presentation of the English translation for the Finnish word that would appear in the upcoming sentence (or a series of dashes for buffer sentences). As soon as participants felt that they had familiarized themselves with the English translation, they pressed the space bar to proceed, at which point the English word was replaced with a fixation point that the participants were to fixate to trigger the appearance of the sentence. This served as a drift correction procedure to adjust for any drift in participants' gaze. As soon as the participant's gaze coincided with the fixation point, the sentence appeared on the screen, with its first letter in the same position where the fixation point had been. The participants pressed the space bar to progress from screen to screen. If a sentence was followed by a comprehension question, they were to press the yes or no key to respond. Figure 2 shows a sample exposure-phase trial sequence with a comprehension question.

Upon completion of the study phase, participants spent 15 min engaged in the distractor task. Then they performed the immediate form recognition and form–meaning mapping posttests, in order. The posttests were untimed. After the completion of the immediate posttests, participants filled out the language background questionnaire. Delayed posttests were held 44–78 hr after the first session and were identical in procedure to the immediate posttests. Participants were not aware beforehand of the content of this second session. The entire experiment took approximately 2 hr to complete.

Analyses and Results

SPSS version 25 (IBM Corp., 2017) was used for all statistical analyses in this study. SPSS version 25, Excel 2013, and PowerPoint 2013 were used for the graphics. Response accuracy on the comprehension questions was acceptable for all participants (82%–100%, $M = 96\%$, $SD = 3.4\%$). There was no significant difference in accuracy between questions that followed massed and spaced repetitions, $t(39) = 0.598$, $p = .552$. A visual inspection of the eye-fixation data revealed overall good quality; therefore, no manual adjustment was performed. Values shorter than 60 ms were deleted. This resulted in the loss of 14 (0.36%), 5 (0.13%), and 2 (0.05%) cases for first fixation duration, gaze duration, and total reading time, respectively.

Background questionnaire

Responses on the background questionnaire showed that none of the participants had any prior knowledge of Finnish or any other language of the Finnic language family. Further, all participants indicated that none of the target words had struck them as familiar upon initial exposure.

Research Question 1

Research Question 1 investigated whether temporal distribution of repeated exposures affects attentional processing of L2 words. Because participants studied the words intentionally, total reading time was analyzed as the primary measure of interest. This is a measure of all overt attention that a target word received, including any intentional rehearsal. Total reading time is known as the most pedagogically relevant eye-tracking measure and as the measure that is the most highly correlated with learning outcomes (Godfroid et al., 2017). Further, first fixation duration and gaze duration were analyzed as early measures of lexical processing (Rayner, 1998; Rayner et al., 2006). These are informative about initial stages of word reading before any strategic attention allocation processes have an effect. Particularly, first fixation duration is known to operate too quickly to be under voluntary control (Rayner & Pollatsek, 1987). A Bonferroni correction was used to correct for multiple testing ($\alpha = .05/12 = .004$).

Total reading time

The total reading time data were positively skewed and had extreme outliers in the upper tail. Figure 3 presents boxplots that show the variance in these data. Figure 4

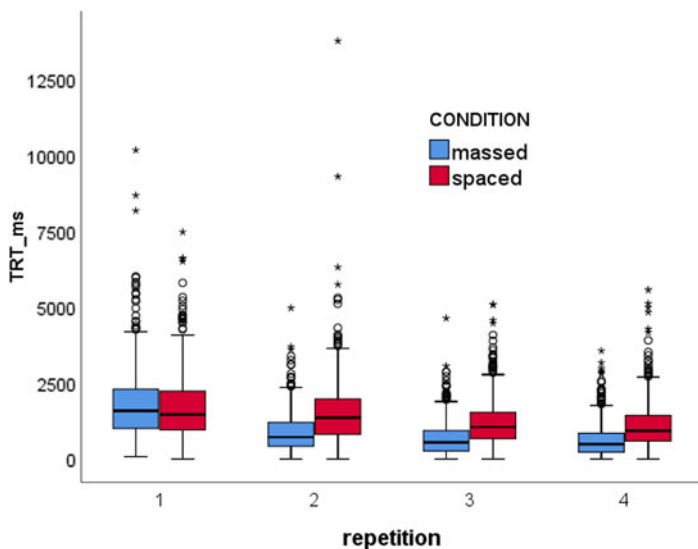


Figure 3. Boxplots showing the variance in total reading time (in ms) at each repetition in the two conditions.

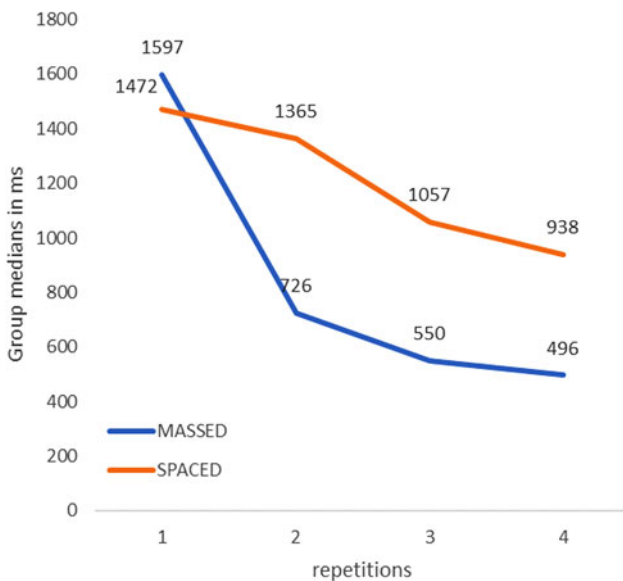


Figure 4. Total reading times in milliseconds at each repetition for the two conditions. This figure shows medians.

presents a line graph (based on medians) that shows how total reading time changed with repeated exposures in the two conditions. It is clear from Figure 4 that, in line with previous findings (Elgort et al., 2017; Godfroid et al., 2017; Joseph et al., 2014; Mohamed, 2017; Pellicer-Sánchez, 2016), reading times decreased with repeated exposures. The graph further shows that this downward trend was steeper in the massed than in the spaced condition, suggesting that the words in the massed condition received less attentional processing across repetitions than the words in the

Table 1. Fit statistics and parameter estimates for multilevel growth-curve models for the rate of change in reading times in the two conditions

		MASSED			SPACED	
		Model 1	Model 2	Model 3	Model 1	Model 2
Fixed components						
Intercept	$\hat{\gamma}_{00}$	28.635***	41.572***	54.672***	35.108***	41.594***
Repetition	$\hat{\gamma}_{10}$		-5.142***	-18.170***		-2.584***
Repetition × Repetition	$\hat{\gamma}_{20}$			2.600***		
Variance of random components						
	$\hat{\lambda}_{700}$	27.708***	29.721***	30.184***	29.209***	29.993***
	$\hat{\sigma}^2$	119.449	85.944	79.029	93.021	84.505
-2LL		14502.482	13895.166	13740.384	14018.886	13842.297

*** $p < .001$.

spaced condition. The downward trend in the massed condition looks a bit curved as if the decrease in attentional processing became less dramatic at later repetitions.

For inferential statistical analyses, observations with residual values greater than 3 *SD* from the mean were trimmed (resulting in the loss of 1.5% of the data) and a square root transformation was performed. A mixed-effects growth curve modeling analysis was used to investigate change processes in reading times across repetitions in the two conditions. A mixed-effects framework was adopted to account for the nested structure of the data, as here each participant contributed multiple data points. The intraclass correlation coefficient (ICC) for the effect of participant was .192, which indicates that 19% of the variability in reading times can be attributed to the differences between participants (Hayes, 2006). The ICC for the effect of items was much smaller (.048). The addition of items as a random effect did not improve and even negatively affected model fit; it also interfered with the convergence of some of the models. Because of these considerations, only participants were used as random effects. Further, the addition of random slopes interfered with the convergence of some of the models; only random intercepts were used for simplicity and consistency. The likelihood ratio test was used to assess improvement in model fit with the addition of new parameters.

An initial model building process showed significant interactions between the effects of condition and both a linear and a quadratic term for repetition, $\chi^2(1) = 89.064$ and 61.413 , respectively, $ps < .001$, suggesting that the two conditions differed in the shape of their trajectories. For this reason, a growth curve model was fit separately for each condition. Table 1 presents fit statistics and parameter estimates for this model building process. In both conditions, the addition of slope significantly improved model fit, $\chi^2(1) = 607.316$, $p < .001$ for the massed condition and $\chi^2(1) = 176.589$, $p < .001$ for the spaced condition, suggesting that reading times did not remain constant across repetitions in the two conditions but rather there was a change present in both. The best fitting model for the massed condition

was found to be a model with quadratic growth—improvement of model fit with the addition of a quadratic term was $\chi^2(1) = 154.783, p < .001$ —confirming the visual impression from Figure 4 of a curve in this trend. For the spaced condition, a linear model showed the best fit—improvement of model fit with the addition of a quadratic term was only $\chi^2(1) = .037, p = .847$ —which suggests that the rate of change in this condition was more or less constant within the four repetitions.

To explore differences in reading times in the two conditions at each repetition, a separate mixed-effects analysis was performed for each repetition with condition as the independent variable. Table 2 shows the results of this analysis. Model fit did not improve with the addition of condition as the independent variable for the first repetition, $\chi^2(1) = 1.126, p = .289$. This makes sense as at this repetition the two conditions do not yet differ. For each subsequent repetition, however, model fit improved significantly with the addition of condition as the independent variable, $\chi^2(1) = 208.832, p < .001$; $\chi^2(1) = 224.687, p < .001$; $\chi^2(1) = 181.860, p < .001$, for repetitions two, three, and four, respectively. Thus, at repetitions two, three, and four, reading times were significantly longer for the words in the spaced than in the massed condition, suggesting that the target words in the spaced condition received greater attentional processing. This confirms statistically the visual impression from Figure 4 of a more dramatic decrease in total reading times across repetitions in the massed condition. Cohen's *d* effect sizes for the four repetitions were $-0.074, 0.619, 0.654, \text{ and } 0.538$, in order. For repetitions two, three, and four, these are moderate-size effects (Cohen, 1988).

Further analyses

Table 3 shows descriptive statistics for first fixation duration and gaze duration. Similarly to total reading time, these two measures exhibit a decrease across the repetitions, with a more dramatic decrease in the massed condition. For statistical analyses, residual values greater than 3 *SD* from the mean were trimmed (resulting in the loss of 1.3% and 1.6% of the data for first fixation duration and gaze duration, respectively) and a square root transformation was performed. As with total reading times, significant interactions were found between the effects of condition and the linear and quadratic terms for repetition for both first fixation duration, $\chi^2(1) = 7.207, p = .007$ and $\chi^2(1) = 6.200, p = .013$, respectively, and gaze duration, $\chi^2(1) = 74.366$ and 59.586 , respectively, $ps < .001$. For this reason, a growth curve model was fit separately in each condition for each measure. For first fixation duration, the addition of slope significantly improved model fit only in the massed condition, $\chi^2(1) = 31.757, p < .001$. While there was, numerically, a downward trend in the spaced condition, it failed to reach statistical significance, $\chi^2(1) = 3.821, p = .051$. The addition of a quadratic term did not significantly improve model fit in the massed condition, $\chi^2(1) = 1.484, p = .223$, unlike what was observed for total reading time. For gaze duration, the addition of slope significantly improved model fit in both conditions, $\chi^2(1) = 272.620, p < .001$ for the massed and $\chi^2(1) = 12.524, p < .001$ for the spaced condition. The addition of a quadratic term improved fit in the massed condition, $\chi^2(1) = 62.323, p < .001$, but not in the spaced condition, $\chi^2(1) = .377, p = .539$, similarly to the pattern observed for total reading time. To investigate differences in first fixation duration

Table 2. Fit statistics and parameter estimates from a mixed-effects linear modeling analysis of the effects of condition on reading times at each repetition

		Repetition 1		Repetition 2		Repetition 3		Repetition 4	
		Model 1	Model 2	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
Fixed components									
Intercept	$\hat{\gamma}_{00}$	39.250***	40.196***	32.465***	19.040***	28.791***	15.147***	27.355***	15.331***
Condition	$\hat{\gamma}_{10}$		-0.628		8.968***		9.107***		8.023***
Variance of random components									
	$\hat{\sigma}_{\gamma_{00}}^2$	52.284***	52.364***	31.598***	32.946***	24.699***	25.841***	23.030***	23.862***
	$\hat{\sigma}_{\gamma_{10}}^2$	81.136	81.028	102.071	81.054	99.186	77.521	92.707	75.898
-2LL		6808.953	6807.827	7160.231	6951.399	7162.872	6938.186	7076.207	6894.347

*** $p < .001$.

Table 3. Descriptive statistics for first fixation duration and gaze duration across repetitions in the two conditions

	Massed repetitions			Spaced repetitions		
	Mean	SD	Median	Mean	SD	Median
First fixation duration						
Repetition 1	270	180	223	275	148	237
Repetition 2	241	122	218	278	167	233
Repetition 3	221	106	206	271	164	230
Repetition 4	214	97	206	260	139	228
Gaze duration						
Repetition 1	792	667	626	764	614	590
Repetition 2	439	379	337	728	544	595
Repetition 3	340	268	261	642	466	543
Repetition 4	326	265	237	660	491	554

and gaze duration between the two conditions, separate analyses were performed at each repetition. The addition of condition as an independent variable made a significant improvement to model fit at each repetition except the first for both first fixation duration, $\chi^2(1) = 3.031, p = .082; 16.151, p < .001; 22.465, p < .001; \text{ and } 28.061, p < .001$ for repetitions one, two, three, and four, respectively; and gaze duration, $\chi^2(1) = 0.556, p = .456; 126.238, p < .001; 169.992, p < .001; \text{ and } 192.709, p < .001$, in the same order. Thus, the results of the first fixation duration and gaze duration analyses also show significant differences between the two conditions, suggesting less attention in the massed than in the spaced condition, in a similar pattern to total reading time. This suggests that, in this intentional learning context, effects of spacing were not limited to intentional allocation of rehearsal but extended to the less voluntary and less strategic processes involved in initial stages of word reading.

A number of other eye-tracking measures were examined descriptively to explore any qualitative differences in attention patterns that may indicate use of different learning strategies. Table 4 presents descriptive statistics on these indices in the two conditions across the four repetitions. It also gives this information for initial exposures (in the massed condition) that occurred evenly spread across the four experimental blocks. These serve as a baseline representing how an initial encounter with a heretofore unseen L2 word was processed attentively at the different time points across the study phase and thus allowed me to investigate any effects of serial position. Fixation count is the total number of times a target word was fixated within a sentence. Skipping rate is the percentage of times a word was not fixated at all within a sentence. First-pass skipping rate is the percentage of times a word was initially skipped but returned to after examining other parts of the sentence. Percent regressions in is a measure of how often a word was revisited after examining other parts of the sentence, regardless of whether or not it had been skipped upon initial pass. Revisit time is the amount of time, in milliseconds, that

Table 4. Descriptive statistics for additional eye-tracking indices for the four repetitions in the two conditions and for initial exposures across the four blocks

Eye-tracking measures	Massed repetitions				Spaced repetitions				First exposures across the blocks				
	<i>M</i>	<i>SD</i>	Mdn	Total Number ^a	<i>M</i>	<i>SD</i>	Mdn	Total Number ^a	<i>M</i>	<i>SD</i>	Mdn	Total Number ^a	
Skipping rate (%)													
Rep 1	0	0	0	0.4	1.8	0	2	Block 1	0	0	0	0	
Rep 2	1.5	3.2	0	7	0.2	1.3	0	1	Block 2	0	0	0	
Rep 3	2.5	4.3	0	12	0.2	1.3	0	1	Block 3	0	0	0	
Rep 4	2.3	4.6	0	11	0.6	2.2	0	3	Block 4	0	0	0	
First-pass skipping rate (%)													
Rep 1	1.5	3.2	0	7	1.5	5.6	0	7	Block 1	0.8	5.3	0	1
Rep 2	3.3	5.3	0	16	1.0	2.8	0	5	Block 2	2.5	8.9	0	3
Rep 3	3.3	6.5	0	16	2.3	4.2	0	11	Block 3	0.8	5.3	0	1
Rep 4	4.2	6.8	0	20	1.9	4.0	0	9	Block 4	1.7	7.4	0	2
Percent regressions in (%)													
Rep 1	80.7	16.1	83.3		81.7	16.0	83.3		Block 1	84.2	22.6	100	
Rep 2	64.3	22.5	66.7		77.5	20.9	83.3		Block 2	80.8	24.9	100	
Rep 3	58.5	24.6	58.3		68.2	24.0	73.9		Block 3	83.8	25.2	100	
Rep 4	55.1	26.8	58.3		60.7	24.7	66.7		Block 4	74.2	27.7	66.7	

(Continued)

Table 4. (Continued)

Eye-tracking measures	Massed repetitions				Spaced repetitions				First exposures across the blocks			
	<i>M</i>	<i>SD</i>	Mdn	Total Number ^a	<i>M</i>	<i>SD</i>	Mdn	Total Number ^a	<i>M</i>	<i>SD</i>	Mdn	Total Number ^a
Fixation count												
Rep 1	4.1	3.1	3		4.0	2.9	3		Block 1	4.4	3.2	3
Rep 2	2.9	1.7	2		3.7	2.5	3		Block 2	4.1	3.2	3
Rep 3	2.8	1.4	2		3.1	2.0	3		Block 3	4.2	3.3	3
Rep 4	2.6	1.4	2		2.9	2.1	2		Block 4	3.7	3.0	3
Revisit time (ms)												
Rep 1	1033	1077	739		971	1019	693		Block 1	1032	953	791
Rep 2	479	598	270		843	1101	605		Block 2	1004	999	708
Rep 3	367	516	181		608	716	404		Block 3	1059	1196	796
Rep 4	334	508	144		470	654	251		Block 4	1037	1157	676
Revisit/TRT ^b (ms)												
Rep 1	0.48	0.33	0.53		0.49	0.32	0.52		Block 1	0.51	0.31	0.59
Rep 2	0.40	0.34	0.43		0.45	0.32	0.49		Block 2	0.48	0.33	0.52
Rep 3	0.36	0.34	0.37		0.40	0.33	0.45		Block 3	0.48	0.31	0.49
Rep 4	0.35	0.35	0.33		0.34	0.33	0.31		Block 4	0.47	0.35	0.52

Note: ^aThis is the total number summed across participants. ^bTRT, total reading time.

participants spent on a target word after examining other parts of the sentence. Revisit/total reading time, gives this amount as a proportion of total reading time.

The patterns in these measures converge in suggesting an overall tendency to give less attention to the target words across repetitions by fixating them fewer times, skipping them more, and revisiting them less often and for shorter durations. Further, this trend is, again, less dramatic in the spaced condition, where words seem to have been processed more attentively overall, on all these measures. These results indicate that there is not much evidence of attention declining across the blocks of the experiment, suggesting that the reduction observed across repetitions is due to a repetition effect and not to fatigue or any changes in strategies as the experiment progressed. The patterns further show that much of total reading time was due to revisiting (revisit/total reading time), particularly in early exposures. This is to be expected as the participants studied the words intentionally. Overall, there were few instances of skipping a word entirely, with this number going up in the massed condition with repetition. There is no evidence of skipping rate going up as the experiment progressed, as the total number remains at zero for initial exposures across the experimental blocks. In sum, no unusual rehearsal patterns were found in these additional measures.

Reminding/study-phase retrieval

In the present study, participants encountered L2 words in sentence contexts that differed from repetition to repetition. Recall that context variability was previously shown to negatively impact learning of spaced repetitions (Johnston & Uhl, 1976; Verhoeven et al., 2004). This is attributed to lower reminding potential of such encounters (Benjamin & Tullis, 2010). It was important to investigate whether a failure of study-phase retrieval may pose a potential threat to learning in the spaced condition. Reading times in the spaced condition showed a significant downward trend across repeated exposures (on gaze duration and total reading time though not on first fixation duration). Such facilitation is usually taken to indicate increased familiarity that comes with repetition (Joseph et al., 2014; Pellicer-Sánchez, 2016; Rayner & Duffy, 1986; Rayner, Raney, & Pollatsek, 1995). Such a repetition effect would, in turn, suggest that the knowledge gained during previous exposures was likely retrieved upon seeing a word again and facilitated processing. One potential confound, however, is that the reduction in reading times could be simply due to a general speedup in reading as the experiment progressed (recall that repetitions in the spaced condition occurred across the blocks and coincided with the number of each block). To isolate true repetition effects, I again used first encounters in the massed condition as a baseline. Recall that first encounters with words in the massed condition occurred across the four blocks. They are, therefore, informative about how much reading time a heretofore unseen word received at the different time points throughout the study phase. For this reason, they provide a useful baseline for isolating true effects of repetition from order effects or any effects of fatigue. A linear mixed-effects model was run separately (due to unequal *n* sizes) for repeated encounters in the spaced condition and baseline first encounters in the massed condition with experimental block as a factor to explore such facilitation more closely. This was done for total reading time, gaze duration, and first fixation duration.

Table 5. Reading times for repeated spaced exposures and first exposures across the four blocks; the table also shows significance values for the differences between the reference (first) block and each subsequent block

Eye-tracking measures	First exposures				Repeated spaced exposures			
	<i>M</i>	<i>SD</i>	Median	Sig. ^a	<i>M</i>	<i>SD</i>	Median	Sig. ^a
Total reading time								
Block 1	1734	956	1615		1727	1118	1472	
Block 2	1810	1166	1646	.828	1561	1145	1365	.001*
Block 3	1853	1440	1560	.570	1243	799	1057	<.001*
Block 4	1869	1265	1634	.697	1127	785	938	<.001*
First fixation duration								
Block 1	276	182	237		275	148	237	
Block 2	281	212	215	.438	278	167	233	.985
Block 3	251	132	230	.428	271	164	230	.149
Block 4	274	187	221	.444	260	139	228	.076
Gaze duration								
Block 1	702	470	631		764	614	590	
Block 2	806	785	595	.648	728	544	595	.615
Block 3	828	698	641	.835	642	466	543	.002*
Block 4	832	676	655	.600	660	491	554	.005

Note: ^a*p* value for the difference between each block and the first (reference) block. *Significant at the Bonferroni corrected $\alpha = .003$.

Here, reading times in Block 1 were compared to reading times in each subsequent block. The same data cleaning procedures were used as previously. The Bonferroni correction was applied ($\alpha = .05/18 = .003$). Table 5 shows descriptive statistics and significance values for the differences between the reference (first) block and each subsequent block. For total reading time, the addition of the independent variable significantly improved model fit in the spaced condition, $\chi^2(1) = 183.190$, $p < .001$, but not in baseline exposures, $\chi^2(1) = 0.962$, $p = .327$. Further, each repeated encounter was processed significantly more quickly than the initial encounter in the spaced condition while initial massed condition encounters in Block 1 did not differ from initial encounters in any of the subsequent blocks (even if the Bonferroni correction were not applied). Because no significant speedup was found on initial encounters across the study phase, an overall speedup can be ruled out as an explanation for the facilitation observed across repetitions in the spaced condition, suggesting a true repetition effect in the spaced condition.

For gaze duration, the addition of block as an independent variable also significantly improved model fit in the spaced condition, $\chi^2(1) = 14.698$, $p < .001$, but not in the baseline, $\chi^2(1) = 0.996$, $p = .318$. As in total reading time, no significant differences were observed in this measure between initial exposures in Block 1 and initial exposures in each subsequent block, suggesting, again, no overall speedup

across time in gaze duration. For the spaced condition, Repetition 2 did not differ significantly from Repetition 1; however, Repetitions 3 and 4 straddled significance at the Bonferroni corrected alpha level. Thus, a repetition effect was observed in later repetitions for gaze duration.

Finally, for first fixation duration, the addition of block as an independent variable did not significantly improve model fit in either condition: the massed condition, $\chi^2(1) = 0.912$, $p = .340$; the spaced condition, $\chi^2(1) = 5.338$, $p = .021$.

While no overall speedup in reading times was observed across the study phase on any of the measures, a repetition effect was quite clearly present. This effect was strongest in total reading time, evident only in later repetitions in gaze duration, and only numerically present in first fixation duration, suggesting that recognition of spaced repetitions had not reached automaticity within the four exposures but required more time, particularly with earlier repetitions.

Research Question 2

Research Question 2 investigated whether intentional learning of L2 words in different sentence contexts is affected by how closely together or widely apart repeated exposures occur. Summed correct responses on immediate and delayed form and meaning tests were used to answer this question. While these were count data, the scores in each test presented a nearly normal distribution; therefore, the data were treated as continuous for the analyses. One participant of 40 did not come back for the delayed posttests. The data were treated as missing at random. Cronbach's α reliability coefficients for the posttests were as follows: form immediate: $\alpha = .790$; form delayed: $\alpha = .776$; meaning immediate: $\alpha = .868$; meaning delayed: $\alpha = .897$. Because the serial position of Repetition 4 for each word could not be equated across the two conditions, binary logistic regression analyses were conducted with serial position as the independent variable and the scores on each of the four posttests as the dependent variables. Serial position was found not to have a significant effect on any of the posttest scores: form immediate, Wald $\chi^2(1) = 1.305$, $p = .254$; form delayed, Wald $\chi^2(1) = 1.238$, $p = .266$; meaning immediate, Wald $\chi^2(1) = 0.034$, $p = .854$; meaning delayed, Wald $\chi^2(1) = 0.885$, $p = .347$.

Table 6 presents overall scores achieved on each of the four tests, averaged across participants. It is clear from this table that there was a wide range of scores. Figures 5 and 6 show boxplots for the scores on the four tests in the two conditions. A visual

Table 6. Total scores on the immediate and delayed form recognition and form-meaning mapping tests; these scores are collapsed across the two conditions and are out of 24 possible points

Posttest	Minimum	Maximum	Mean	SD
Form immediate	5	23	14.9	4.64
Form delayed	6	23	14.7	4.66
Meaning immediate	0	20	10	5.83
Meaning delayed	0	20	8.9	6.26

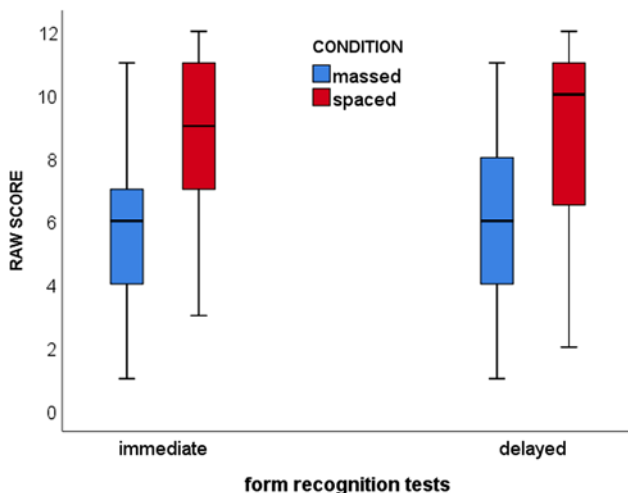


Figure 5. Immediate and delayed form recognition posttest scores. These scores are out of 12 possible points and represent the sum of correct answers in each condition, averaged across participants.

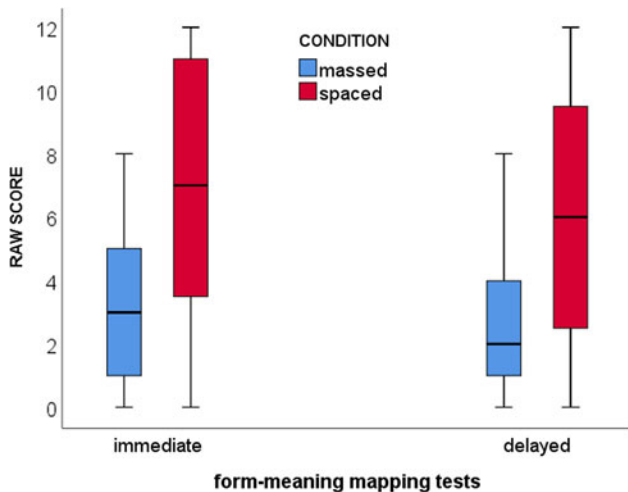


Figure 6. Immediate and delayed form-meaning mapping posttest scores. These scores are out of 12 possible points and represent the sum of correct answers in each condition, averaged across participants.

inspection of the boxplots suggests that the scores were higher in the spaced condition on all four tests and also that the scores on the form recognition test were higher than those on the form-meaning mapping test. It is also clear that there was not much change in the scores across time.

A linear mixed-effects analysis exploring the effects of condition (massed, spaced), test type (form, meaning), and retention interval (immediate, delayed) as the independent variables was performed on the raw scores, with participants as random effects. The ICC for the effect of participant was .379, which shows how variable the responses were across participants. This analysis showed significantly higher scores in the spaced relative to the massed condition, $\chi^2(1) = 126.803$, $p < .001$, and significantly higher scores on the form recognition test than on the

form–meaning mapping test, $\chi^2(1) = 139.757$, $p < .001$, but no effect of retention interval, $\chi^2(1) = 2.389$, $p = .122$. The addition of interaction terms produced the following results, in order: retention interval by condition, $\chi^2(1) = 2.152$, $p = .142$; condition by test type, $\chi^2(1) = 5.763$, $p = .016$; retention interval by test type, $\chi^2(1) = 1.483$, $p = .223$. The three-way interaction was not significant, $\chi^2(1) = 0.085$, $p = .770$. Note that there is no significant interaction between retention interval and condition. Note, however, that the interaction between condition and test type is significant. This confirms the visual impression from Figures 5 and 6 of a greater difference in scores between the two conditions on the form–meaning mapping test than on the form recognition test. The significant interaction also means that the effect of test type (performance on the form test being higher overall than on the meaning test) is larger in the massed than in the spaced condition. A linear mixed-effects modeling analysis was further performed for each of the four tests separately with participants as the random effects and condition as the fixed effect. Only random intercepts were used. The Bonferroni correction was applied to adjust for multiple testing ($\alpha = .05/5 = .01$). The ICC for the effect of participant in the four tests was .26, .14, .28, and .21, respectively, for form immediate, form delayed, meaning immediate, and meaning delayed tests. Table 7 presents the results of this analysis. The addition of condition as the independent variable significantly improved model fit as indicated by the difference in -2LL between the models with and without this independent variable: form immediate, $\chi^2(1) = 37.881$, $p < .001$; form delayed, $\chi^2(1) = 25.971$, $p < .001$; meaning immediate, $\chi^2(1) = 50.545$, $p < .001$; meaning delayed, $\chi^2(1) = 47.310$, $p < .001$. This independent variable was statistically significant ($ps < .001$) for all four tests. Cohen's d effect sizes for the effect of condition in the four tests were as follows: form immediate, 1.24; form delayed, 0.98; meaning immediate, 1.94; and meaning delayed, 1.91. These are large effect sizes (Cohen, 1988).

Research Question 3

Research Question 3 investigated whether the positive effects that spacing repeated encounters with novel L2 vocabulary was found to have on learning operates through increased attentional processing of spaced encounters relative to massed encounters, as predicted by the deficient processing account of the spacing effect. A mediation analysis was used to address this question. In mediation, the effect of a predictor variable on the outcome variable (direct effect) operates through a mediator variable (indirect effect). In the present case, I tested whether the direct effect of spacing on learning (as measured by the posttest scores) is mediated by the indirect effect of attention (inferred from the total reading times).

For the present analysis, scores from the four posttests that measured learning were collapsed into one score per condition per participant. Recall that the same basic pattern of results was found across the four posttests. Table 8 shows Pearson correlations among the four tests. These correlations are quite high. Table 8 also presents the results of a principle component analysis, which showed that the four tests load quite highly and quite uniformly on one underlying component. This suggests that the four tests are likely measuring the same underlying construct. In such a case, the scores can be summed or averaged without much loss

Table 7. Parameter estimates from a mixed-effects linear modeling analysis of the effects of condition on learning gains as measured by the four posttests

		Form immediate		Form delayed		Meaning immediate		Meaning delayed	
		Model 1	Model 2	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
Fixed components									
Intercept	$\hat{\gamma}_{00}$	7.425***	2.775***	7.333***	3.718***	5.013***	-0.875	4.449***	-0.744
Condition	$\hat{\gamma}_{10}$		3.10***		2.410***		3.925***		3.462***
Variance of random components									
	$\hat{\sigma}_{\gamma_{00}}$	1.319	3.722**	2.299	3.751**	2.925	6.776***	5.286*	8.281***
	$\hat{\sigma}^2$	7.850	3.045	5.974	3.070	10.738	3.035	8.526	2.535
-2LL		403.463	365.582	383.039	357.068	434.325	383.780	419.967	372.657

* $p < .05$.** $p < .01$.*** $p < .001$.

Table 8. Pearson correlation coefficients and principle components analysis results for the four vocabulary posttests

Posttest	Pearson correlation coefficients				Principle components analysis
	Form immediate	Form delayed	Meaning immediate	Meaning delayed	Component 1 (85% variance explained)
Form immediate	1				.912
Form delayed	.842**	1			.908
Meaning immediate	.775**	.741**	1		.932
Meaning delayed	.758**	.773**	.923**	1	.937

***p* < .01.

of information. The present analysis was performed on averaged scores. The reading times were summed across the four exposures for each participant. This sum represented the total attentional processing a word received in the study phase. Figure 7 presents graphically the conceptual structure of the analysis with standardized coefficients for the three simple effects tested separately, including the direct effect of spacing on learning. The indirect effect, which was of primary interest, was tested with the PROCESS Procedure for SPSS Version 3.0 (Hayes, 2018) using bootstrapped 95% confidence intervals with 10,000 bootstrap samples. The partially standardized effect was .3184, bootstrapped standard error = .0979, 95% bootstrapped confidence interval [.1435, .5324]. The confidence interval for the indirect effect does not include zero, which indicates significant mediation. Thus, attention was found to be a significant mediator for the spacing effect in L2 word learning obtained in this study. This confirms the predictions of the deficient processing account. Recall that this account states that the positive effects of spacing on learning are due to increased attentional processing of spaced repetitions relative to massed repetitions. It can be concluded that spacing of repeated exposures with L2 words positively affects learning through increased attentional processing; that is, spacing has a positive effect on attention, which, in turn, has a positive effect on learning.

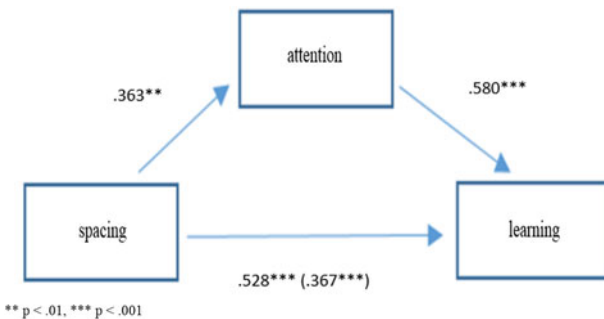


Figure 7. The path diagram for the mediation analysis. Standardized coefficients for each simple effect are presented alongside the corresponding arrow. The standardized coefficient for the effect of spacing on learning, controlling for attention, appears in parentheses. ***p* < .01. ****p* < .001.

Discussion

The present study tested the contribution of deficient processing to the spacing effect in intentional L2 vocabulary learning from sentence contexts. Participants studied 24 novel Finnish words by reading each in four different English sentence contexts. The results showed that words that repeated in sentences that were separated by other sentences and an intervening distractor task (for a total of 18 min, on average, between repetitions) were remembered significantly better than words that repeated in consecutive sentences, as measured by immediate and 48- to 72-hr-delayed form recognition and form–meaning mapping posttests. Thus, a spacing effect was obtained under the present conditions. The effect sizes were quite large, suggesting an advantage for studying words in a distributed fashion that is of practical importance. It is significant that such large effect sizes were obtained with only 18 min, on average, between repetitions, and that a difference of this magnitude can be seen in as few exposures as four.

Results from the analyses of total reading times showed that words that occurred in sentences that were spaced received significantly more attention than words that occurred in consecutive sentences. This suggests that temporal distribution of repeated exposures to L2 words plays a role in how much attentional processing the words receive; namely, repeated exposures that are close in time receive less attention. Analyses of first fixation durations and gaze durations as early measures of lexical processing showed a similar pattern to that found in total reading time: massed repetitions received significantly less attentional processing. This shows that, even though participants engaged in intentional study, effects of spacing extended to the less voluntary processes involved in early stages of word reading before any intentional rehearsal may have an effect. The present results suggest, therefore, that both a controlled (Greene, 1989) and an automatic (Challis, 1993; Toppino, 1991) deficient processing account may be relevant for intentional learning of L2 vocabulary from context. An examination of other eye-tracking indices did not reveal any unusual rehearsal patterns or any differences in eye-movement patterns beyond an overall decrease in attention across repetitions and this decrease being more dramatic in the massed condition.

In the present study, L2 words were encountered in contexts that differed from repetition to repetition. This is how L2 words are commonly encountered in language learning. However, this may have important consequences for both attention and learning. Such difference in contexts likely precluded deficient processing in its strictest sense (as seen in psychology studies with word list learning) as each repeated encounter with a word had to be processed in a new sentence context. It is, therefore, significant that spacing still had the effect of increasing attention above and beyond the additional processing engendered by such context variability and that this effect was not limited to intentional rehearsal.

Another important consequence of context variability is its effects on learning. Previous research in psychology found that contextual change enhances learning of massed repetitions but impedes learning of spaced repetitions because it may contribute to a failure of study-phase retrieval at longer lags (Verkoeijen et al., 2004). Words in the spaced condition showed facilitation in reading times across repetitions that could not be attributed to effects of serial position and therefore indicated

a repetition effect, suggesting that repeated spaced exposures were processed as repetitions. Thus, repetitions in the spaced condition experienced retrieval that was effortful yet successful. This is believed to be a “sweet spot” for optimal learning in dual mechanism accounts that combine deficient processing with the assumption of study-phase retrieval/reminding (Benjamin & Tullis, 2010). This may account for the large benefits that spacing produced in this study. The fact that memory traces survived between spaced repetitions may be due to the specific combination of the length of the interstudy interval, the fact that learners engaged in intentional learning, which may establish stronger memory traces (Verkoeijen et al., 2005), and the relative ease of the task (Bui, Maddox, & Balota, 2013; Elgort & Warren, 2014; Verkoeijen & Bouwmeester, 2008).

Finally, the present results showed that amount of attentional processing mediates the spacing effect in L2 vocabulary learning, in line with the predictions of the deficient processing theory of the spacing effect (Cuddy & Jacoby, 1982; Zechmeister & Shaughnessy, 1980). The results showed that, even when words are encountered in different contexts, learners process these repeated encounters more attentively if they occur more widely spaced over time, which in turn results in more learning.

Implications for vocabulary learning and teaching

It is widely accepted that attention is important for learning many aspects of an L2, including vocabulary (Gass, 1988; Godfroid et al., 2013; Laufer & Hulstijn, 2001; Mohamed, 2017; Schmidt, 1990, 2001; Schmitt, 2008). Researchers and practitioners alike have been looking for effective ways of inducing more attention to target forms. A number of techniques have been proposed to accomplish this, such as input enhancement (Sharwood Smith, 1993) and the use of the noticing function of output (Swain, 1995), to name just a few. The present results suggest that effective spacing of repeated exposures with L2 forms can be added to this list. In the present study, the length of the interval that separated repeated exposures to novel L2 words had a significant effect on how attentively these exposures were processed. It is not unreasonable to suppose that the finding of more attention to spaced repetitions may generalize to a different time scale and to different learning targets as well as activities that differ from sentence reading; that is, language forms that repeat closely in any kind of input or are focused on in a massed fashion during a lesson may receive less attentional processing. A recommendation can be made to avoid massing repeated exposures or treatments of the same form to keep learners' attention to target forms at a higher level.

While in the present study participants studied the words intentionally, eye-tracking indices of early lexical processing known to be under less voluntary control suggest that the above recommendation may be useful for incidental vocabulary acquisition contexts as well. Thus, even when a learner is not trying to commit a word to memory but only processes it for recognition and comprehension, repeated exposures that are close together may receive less attentional processing and may, therefore, be not as useful for learning as they would be if they were more widely spaced.

While it is unlikely that increasing interstudy interval will ever have the effect of reducing attention to a target form, generalizing learning effects far beyond the time

scale of the present study is less straightforward as learning may be negatively affected when repetitions are spaced too widely (Toppino & Bloom, 2002; Verkoeijen et al., 2005). Optimal interstudy interval may, further, be shorter the more complex the task (Bui et al., 2013; Donovan & Radosevich, 1999; Elgort & Warren, 2014; Suzuki & DeKeyser, 2017; Verkoeijen & Bouwmeester, 2008).

The findings of this study also have implications for discussions of how many exposures are needed to learn a word (Nation, 1990). The present results are in line with prior studies that have shown that, holding constant the number of exposures to an L2 word, more attentional processing leads to more learning (Godfroid et al., 2017; Mohamed, 2017). Thus, if we can induce more attention to L2 words, such as through effective use of spacing, we may be able to achieve more learning with fewer exposures, which may mean time saved.

Limitations and future directions

The present study has limitations that future research needs to address. This was a first attempt at testing attention as a mediator for the spacing effect in contextual L2 vocabulary learning. The design was a departure from the traditional word list rote learning paradigm widely used in psychology to investigate deficient processing as in the present study words were encountered in contexts that differed from repetition to repetition, which approximates more closely the way L2 vocabulary is encountered in language learning but may, as discussed earlier, have important consequences for both attention and learning. Not to make too drastic a departure from established research, the study aimed to keep the learning process simple. Participants studied words denoting simple and generic concepts such as “butterfly,” “food,” and “city” in a language that was completely novel to them. The use of a novel language and such simple words allowed me to target simple processes of word learning characteristic of novice L2 vocabulary learning stages before subtleties in meaning and stylistic and other distributional features of a word may become important. The use of native language sentence contexts made targeting such novice learning possible while allowing the use of more interesting sentence contexts. Using sentence contexts allowed to control the spacing of repetitions and to use a counter-balanced design. While these design features allowed me to isolate the effects of interest with more certainty, this level of control creates a limitation because it does not capture all processes involved in authentic L2 reading. Future research needs to investigate these processes in more ecologically valid designs by presenting target words within L2 reading contexts, such as extended L2 texts. Important variables whose effects need to be systematically investigated with L2 reading include learner proficiency and task complexity (Bui et al., 2013; Verkoeijen & Bouwmeester, 2008). Some previous failures to observe a spacing effect in SLA have been attributed to task complexity (Suzuki & DeKeyser, 2017) or lower proficiency and probable memory trace decay from repetition to repetition (Elgort & Warren, 2014), which is consistent with a failure of study-phase retrieval. These studies further used longer interstudy intervals (e.g., mostly 24 hr in Elgort & Warren, who investigated incidental learning of vocabulary from reading). By manipulating interstudy interval on the one hand and task difficulty and learner proficiency on the other, future research may observe important interactions. There may even be a situation in which more

attention may actually not lead to more learning from repetition (contrary to the consistent finding) because of a failure of study-phase retrieval at longer interstudy intervals.

Future research also needs to explore how temporal distribution affects both attention and learning beyond four exposures. In the present study, the rate of change in reading times had a quadratic shape in the massed condition and a linear shape in the spaced condition. It is likely that the spaced condition will also show a trend that deviates from linearity at a point beyond four repetitions. At Repetition 4 there was still a significant difference in reading times between the two conditions. With more repetitions, there may come a point at which the two lines converge and there is no longer a difference. This point, in turn, will likely depend on the interstudy interval and a combination of other moderator variables in any given learning situation. Finally, the benefits that come from each additional exposure need to be investigated as a function of interstudy interval and other relevant variables (see, e.g., Maddox & Balota, 2015).

The present findings tell us that if vocabulary is studied repeatedly in a massed fashion, there may come a point at which these repetitions no longer engage as much attention and are no longer as useful for learning as they would be if the words were, instead, to be revisited at a later time. To be able to give more precise recommendations for teaching/learning practice and for materials and syllabus design, more research is needed that investigates systematically what the optimal spacing may be for the acquisition of vocabulary and of other aspects of L2. The optimal interstudy interval will, in turn, depend on other variables present in any given L2 learning situation. These include whether learning is intentional/incidental, the presence/absence of feedback, individual learner differences such as proficiency and working memory capacity, whether the learners are engaged in comprehending input or producing output, and the input/output modality, to name just a few. Future research should not neglect to investigate the underlying mechanisms responsible for the spacing effect in SLA contexts. An understanding of the underlying cognitive operations that are involved will allow us to better understand how and why certain variables that are present in our learning contexts may exert a moderating influence on the relationship between spacing and learning. This, in turn, will allow us to set up inquiries in more useful ways such that they do a better job of informing L2 pedagogy about how best to make use of the memory phenomenon that is the spacing effect.

Appendix A. Target Words

List A

aviomies (husband)
 silta (bridge)
 tarina (story)
 savuke (cigarette)
 puhelin (phone)
 toimisto (office)
 kieli (language)
 valmis (ready)

kaupunki (city)
taivas (sky)
keltainen (yellow)
lyhyt (short)

List B

avain (key)
perhonen (butterfly)
elokuva (movie)
paita (shirt)
rakennus (building)
lattia (floor)
kirjasto (library)
osoite (address)
kasvot (face)
ruoka (food)
valkoinen (white)
laukku (bag)

Appendix B. Sample Sentences

A sample experimental sentence with a comprehension question

The article was written in a kieli that he did not know.

- Could he read and understand the article?

A sample buffer trial with a comprehension question

Her luggage was so heavy that she had to ask for help.

- Was she probably traveling?

Author ORCID*s*.  Natalie Koval, [0000-0001-9233-0717](https://orcid.org/0000-0001-9233-0717)

Acknowledgments. This research was supported in part by the Second Language Studies Doctoral Program at Michigan State University. I would like to thank Drs. Patti Spinner and Aline Godfroid for their valuable feedback on the experimental design and data analysis. I am also grateful to MSU CSTAT statistical consulting for feedback on the statistical analyses. I am further grateful to associate editor Dr. Annie Tremblay and my three anonymous reviewers for their valuable comments and suggestions that helped me to greatly improve my contribution.

References

- Bahr*ick*, H. P., Bahr*ick*, L. E., Bahr*ick*, A. S., & Bahr*ick*, P. E. (1993). Maintenance of foreign language vocabulary and the spacing effect. *Psychological Science*, *4*, 316–321.
- Beg*g*, I., & Gre*en*, C. (1988). Repetition and trace interaction: Superadditivity. *Memory & Cognition*, *16*, 232–242.

- Bellezza, F. S., Winkler, H. B., & Andrasik, F. (1975). Encoding processes and the spacing effect. *Memory & Cognition*, *3*, 451–457.
- Benjamin, A. S., & Tullis, J. G. (2010). What makes distributed practice effective? *Cognitive Psychology*, *61*, 228–247.
- Bird, S. (2010). Effects of distributed practice on the acquisition of second language English syntax. *Applied Psycholinguistics*, *31*, 635–650.
- Bjork, R. A., & Allen, T. W. (1970). The spacing effect: Consolidation or differential encoding? *Journal of Verbal Learning and Verbal Behavior*, *9*, 567–572.
- Bloom, K. C., & Shuell, T. J. (1981). Effects of massed and distributed practice on the learning and retention of second-language vocabulary. *Journal of Educational Research*, *74*, 245–248.
- Bower, G. H. (1972). Stimulus-sampling theory of encoding variability. In A. W. Melton, and E. Martin (Eds.), *Coding processes in human memory* (pp. 85–123). Washington, DC: Winston.
- Brybaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition*, *1*, 1–20.
- Bui, D. C., Maddox, G. B., & Balota, D. A. (2013). The roles of working memory and intervening task difficulty in determining the benefits of repetition. *Psychonomic Bulletin & Review*, *20*, 341–347.
- Callan, D., & Schweighofer, N. (2010). Neural correlates of the spacing effect in explicit verbal semantic encoding support the deficient-processing theory. *Human Brain Mapping*, *31*, 645–659.
- 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*, 354–380.
- Cepeda, N. J., Vul, E., Rohrer, D., Wixted, J. T., & Pashler, H. (2008). Spacing effects in learning: A temporal ridge of optimal retention. *Psychological Science*, *19*, 1095–1102.
- Challis, B. H. (1993). Spacing effects on cued-memory tests depend on level of processing. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *19*, 389–396.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Earlbaum.
- Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*, 671–684.
- Cuddy, L. J., & Jacoby, L. L. (1982). When forgetting helps memory: An analysis of repetition effects. *Journal of Verbal Learning and Verbal Behavior*, *21*, 451–467.
- Davies, M. (2008–). The Corpus of Contemporary American English (COCA): 520 million words, 1990–present. Available online at <https://corpus.byu.edu/coca/>
- Delaney, P. F., Verkoeijen, P., & Spigler, A. (2010). Spacing and testing effects: A deeply critical, lengthy, and at times discursive review of the literature. In B. H. Ross (Ed.), *Psychology of learning and motivation: Advances in research and theory* (pp. 63–147). San Diego: Elsevier Academic Press.
- Dellarosa, D., & Bourne, L. E. (1985). Surface form and the spacing effect. *Memory & Cognition*, *13*, 529–537.
- Dempster, F. N. (1988). The spacing effect: A case study in the failure to apply the results of psychological research. *American Psychologist*, *43*, 627–634.
- Donovan, J. J., & Radosevich, D. J. (1999). A meta-analytic review of the distribution of practice effect: Now you see it, now you don't. *Journal of Applied Psychology*, *84*, 795–805.
- Ebbinghaus, H. (1885). *Über das gedächtnis: Untersuchungen zur experimentellen psychologie*. Berlin: Duncker & Humblot.
- Elgort, I., Brybaert, M., Stevens, M., & Van Assche, E. (2017). Contextual word learning during reading in a second language: An eye-movement study. *Studies in Second Language Acquisition*. Advance online publication.
- Elgort, I., & Warren, P. (2014). L2 Vocabulary learning from reading: Explicit and tacit lexical knowledge and the role of learner and item variables. *Language Learning*, *64*, 365–414.
- Estes, W. K. (1955). Statistical theory of distributional phenomena in learning. *Psychological Review*, *62*, 369–377.
- Gass, S. (1988). Integrating research areas: A framework for second language studies. *Applied Linguistics*, *9*, 198–217.
- Gerbier, E., & Toppino, T. C. (2015). The effect of distributed practice: Neuroscience, cognition, and education. *Trends in Neuroscience and Education*, *4*, 49–59.
- Glenberg, A. M. (1976). Monotonic and nonmonotonic lag effects in paired-associate and recognition memory paradigms. *Journal of Verbal Learning and Verbal Behavior*, *15*, 1–16.

- Glenberg, A. M.** (1979). Component-levels theory of the effects of spacing of repetitions on recall and recognition. *Memory & Cognition*, *7*, 95–112.
- Glenberg, A. M., & Smith, S. M.** (1981). Spacing repetitions and solving problems are not the same. *Journal of Verbal Learning and Verbal Behavior*, *20*, 110–119.
- Godfroid, A., Ahn, J., Choi, I., Ballard, L., Cui, Y., Johnston, S., . . . Yoon, H-J.** (2017). Incidental vocabulary learning in a natural reading context: An eye-tracking study. *Bilingualism: Language and Cognition*. Advance online publication.
- Godfroid, A., Boers, F., & Housen, A.** (2013). An eye for words: Gauging the role of attention in incidental L2 vocabulary acquisition by means of eye tracking. *Studies in Second Language Acquisition*, *35*, 483–517.
- Godfroid, A., & Schmidtke, J.** (2013). What do eye movements tell us about awareness? A triangulation of eye-movement data, verbal reports and vocabulary learning scores. In J. M. Bergsleithner, S. N. Frota, and J. K. Yoshioka (Eds.), *Noticing and second language acquisition: Studies in honor of Richard Schmidt* (pp. 183–205). Honolulu, HI: University of Hawai'i, National Foreign Language Resource Center.
- Greene, R. L.** (1989). Spacing effects in memory: Evidence for a two-process account. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 371–377.
- Greeno, J. G.** (1967). Paired-associate learning with short-term retention: Mathematical analysis and data regarding identification of parameters. *Journal of Mathematical Psychology*, *4*, 430–472.
- Hayes, A. F.** (2006). A primer on multilevel modeling. *Human Communication Research*, *32*, 385–410.
- Hayes, A. F.** (2018). *Introduction to mediation, moderation, and conditional process analysis* (2nd ed.). New York: Guilford Press.
- Helsdingen, A., van Gog, T., & van Merriënboer, J.** (2011). The effects of practice schedule and critical thinking prompts on learning and transfer of a complex judgment task. *Journal of Educational Psychology*, *103*, 383–398.
- Hintzman, D. L., & Block, R. A.** (1973). Memory for the spacing of repetitions. *Journal of Experimental Psychology*, *99*, 70.
- Hintzman, D. L., Block, R. A., & Summers, J. J.** (1973). Modality tags and memory for repetitions: Locus of the spacing effect. *Journal of Memory and Language*, *12*, 229–238.
- Horst, M., Cobb, T., & Meara, P.** (1998). Beyond A Clockwork Orange: Acquiring second language vocabulary through reading. *Reading in a Foreign Language*, *11*, 207–223.
- Hyönä, J., & Niemi, P.** (1990). Eye movements in repeated movements of a text. *Acta Psychologica*, *73*, 259–280.
- IBM Corp.** (2017). *IBM SPSS Statistics for Windows, Version 25.0*. Armonk, NY: Author.
- Jacoby, L. L.** (1978). On interpreting the effects of repetition: Solving a problem versus remembering a solution. *Journal of Verbal Learning and Verbal Behavior*, *17*, 649–667.
- Johnston, W. A., & Uhl, C. N.** (1976). The contributions of encoding effort and variability to the spacing effect on free recall. *Journal of Experimental Psychology: Human Learning and Memory*, *2*, 153–160.
- Joseph, H., Wonnacott, E., Forbes, P., & Nation, K.** (2014). Becoming a written word: Eye-movements reveal order of acquisition effects following incidental exposure to new words during silent reading. *Cognition*, *133*, 238–248.
- Just, M. A., & Carpenter, P. A.** (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, *87*, 329–354.
- Kang, S., Lindsey, R. V., Mozer, M. C., & Pashler, H.** (2014). Retrieval practice over the long term: Should spacing be expanding or equal-interval? *Psychonomic Bulletin & Review*, *21*, 1544–1550.
- Kapler, I. V., Weston, T., & Wiseheart, M.** (2015). Spacing in a simulated undergraduate classroom: Long-term benefits for factual and higher-level learning. *Learning and Instruction*, *36*, 38–45.
- Kim, M., Kim, J., & Kwon, J. S.** (2001). The effect of immediate and delayed word repetition on event-related potential in a continuous recognition task. *Cognitive Brain Research*, *11*, 387–396.
- Kornell, N., & Bjork, R. A.** (2008). Learning concepts and categories: Is spacing the “enemy of induction”? *Psychological Science*, *19*, 585–592.
- Krug, D., Davis, B., & Glover, J. A.** (1990). Massed versus distributed repeated reading: A case of forgetting helping recall? *Journal of Educational Psychology*, *82*, 366–371.
- Landauer, T. K.** (1969). Reinforcement as consolidation. *Psychological Review*, *76*, 82–96.
- Laufer, B., & Hulstijn, J. H.** (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics*, *22*, 1–26.

- Maddox, G. B.** (2016). Understanding the underlying mechanism of the spacing effect in verbal learning: A case for encoding variability and study-phase retrieval. *Journal of Cognitive Psychology*, *28*, 684–706.
- Maddox, G. B., & Balota, D. A.** (2015). Retrieval practice and spacing effects in young and older adults: An examination of the benefits of desirable difficulty. *Memory & Cognition*, *43*, 760–774.
- Madigan, S. A.** (1969). Intraserial repetition and coding processes in free recall. *Journal of Verbal Learning and Verbal Behavior*, *8*, 828–835.
- Magliero, A.** (1983). Pupil dilations following pairs of identical and related to-be-remembered words. *Memory & Cognition*, *11*, 609–615.
- Mammarella, N., Avons, S. E., & Russo, R.** (2004). A short-term perceptual priming account of spacing effects in explicit cued-memory tasks for unfamiliar stimuli. *European Journal of Cognitive Psychology*, *16*, 387–402.
- Melton, A. W.** (1967). Repetition and retrieval from memory. *Science*, *158*, 532.
- Melton, A. W.** (1970). The situation with respect to the spacing of repetitions and memory. *Journal of Verbal Learning and Verbal Behavior*, *9*, 596–606.
- Mohamed, A. A.** (2017). Exposure frequency in L2 reading: An eye-movement perspective of incidental vocabulary learning. *Studies in Second Language Acquisition*. Advance online publication.
- Nakata, T.** (2015). Effects of expanding and equal spacing on second language vocabulary learning: Does gradually increasing spacing increase vocabulary learning? *Studies in Second Language Acquisition*, *37*, 677–711.
- 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*, 523–552.
- Nation, I. S. P.** (1990). *Teaching and learning vocabulary*. New York: Newbury House.
- Pashler, H., Zarow, G., & Triplett, B.** (2003). Is temporal spacing of tests helpful even when it inflates error rates? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 1051–1057.
- Pavlik, P. I., & Anderson, J. R.** (2005). Practice and forgetting effect on vocabulary memory: An activation-based model of the spacing effect. *Cognitive Science*, *29*, 559–586.
- Pellicer-Sánchez, A.** (2016). Incidental L2 vocabulary acquisition from and while reading: An eye-tracking study. *Studies in Second Language Acquisition*, *38*, 97–130.
- Raaijmakers, J. G. W.** (2003). Spacing and repetition effects in human memory: Application of the SAM model. *Cognitive Science*, *27*, 431–452.
- Rayner, K.** (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*, 372–422.
- Rayner, K.** (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, *62*, 1457–1506.
- Rayner, K., & Duffy, S. A.** (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, *14*, 191–201.
- Rayner, K., & Pollatsek, A.** (1987). Eye movements in reading: A tutorial review. In M. Coltheart (Ed.), *Attention and performance* (Vol. 12, pp. 327–362). London: Erlbaum.
- Rayner, K., Raney, G. E., & Pollatsek, A.** (1995). Eye movements and discourse processing. In R. F. Lorch, and E. J. O'Brien (Eds.), *Sources of coherence in reading* (pp. 9–36). Hillsdale, NJ: Erlbaum.
- Reder, L. M., & Anderson, J. R.** (1982). Effects of spacing and embellishment on memory for the main points of a text. *Memory & Cognition*, *10*, 97–102.
- Robinson, P.** (2003). Attention and memory during SLA. In C. Doughty, and M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 631–678). Oxford: Blackwell.
- Rogers, J.** (2015). Learning second language syntax under massed and distributed conditions. *TESOL Quarterly*, *49*, 857–866.
- Rohrer, D., & Pashler, H.** (2007). Increasing retention without increasing study time. *Current Directions in Psychological Science*, *16*, 183–186.
- Rose, R. J.** (1984). Processing time for repetitions and the spacing effect. *Canadian Journal of Psychology/Revue Canadienne De Psychologie*, *38*, 537–550.
- Ross, B. H., & Landauer, T. K.** (1978). Memory for at least one of two items: Test and failure of several theories of spacing effects. *Journal of Verbal Learning and Verbal Behavior*, *17*, 669–680.
- Rundus, D.** (1971). Analysis of rehearsal processes in free recall. *Journal of Experimental Psychology*, *89*, 63–77.

- Russo, R., & Mammarella, N. (2002). Spacing effects in recognition memory: When meaning matters. *European Journal of Cognitive Psychology*, *14*, 49–59.
- Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, *11*, 129–158.
- Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). New York: Cambridge University Press.
- Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language Teaching Research*, *12*, 329–363.
- Schuetze, U. (2015). Spacing techniques in second language vocabulary acquisition: Short-term gains vs. long-term memory. *Language Teaching Research*, *19*, 28–42.
- Seabrook, R., Brown, G. D., & Solity, J. E. (2005). Distributed and massed practice: From laboratory to classroom. *Applied Cognitive Psychology*, *19*, 107–122.
- Sharwood Smith, M. (1993). Input enhancement in instructed SLA: Theoretical bases. *Studies in Second Language Acquisition*, *15*, 165–179.
- Shaughnessy, J. J., Zimmerman, J., & Underwood, B. J. (1972). Further evidence on the MP-D effect in free-recall learning. *Journal of Verbal Learning and Verbal Behavior*, *11*, 1–12.
- Sobel, H. S., Cepeda, N. J., & Kapler, I. V. (2011). Spacing effects in real-world classroom vocabulary learning. *Applied Cognitive Psychology*, *25*, 763–767.
- Suzuki, Y., & DeKeyser, R. (2017). Effects of distributed practice on the proceduralization of morphology. *Language Teaching Research*, *21*, 166–188.
- Swain, M. (1995). Three functions of output in second language learning. In G. Cook, and B. Seidlhofer (Eds.), *Principle and practice in applied linguistics: Studies in honor of H. G. Widdowson* (pp. 125–144). Oxford: Oxford University Press.
- 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*, 529–536.
- Toppino, T. C. (1991). The spacing effect in young children's free recall: Support for automatic-process explanations. *Memory & Cognition*, *19*, 159–167.
- 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*, 437–444.
- Van Strien, J. W., Verkoeijen, P. P., Van der Meer, N., & Franken, I. H. A. (2007). Electrophysiological correlates of word repetition spacing: ERP and induced band power old/new effects with massed and spaced repetitions. *International Journal of Psychophysiology*, *66*, 205–214.
- Verkoeijen, P. P., & Bouwmeester, S. (2008). Using latent class modeling to detect bimodality in spacing effect data. *Journal of Memory and Language*, *59*, 545–555.
- Verkoeijen, P. P., & Delaney, P. F. (2008). Rote rehearsal and spacing effects in the free recall of pure and mixed lists. *Journal of Memory and Language*, *58*, 35–47.
- Verkoeijen, P. P., Rikers, R. M., & Schmidt, H. G. (2004). Detrimental influence of contextual change on spacing effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 796–800.
- Verkoeijen, P. P., Rikers, R. M., & Schmidt, H. G. (2005). Limitations to the spacing effect: Demonstration of an inverted U-shaped relationship between interrepetition spacing and free recall. *Experimental Psychology*, *52*, 257–263.
- Vlach, H. A., & Sandhofer, C. M. (2012). Distributing learning over time: The spacing effect in children's acquisition and generalization of science concepts. *Child Development*, *83*, 1137–1144.
- Wahlheim, C. N., Dunlosky, J., & Jacoby, L. L. (2011). Spacing enhances the learning of natural concepts: An investigation of mechanisms, metacognition, and aging. *Memory & Cognition*, *39*, 750–763.
- Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics*, *28*, 46–65.
- Whitten, W. B., & Bjork, R. A. (1977). Learning from tests: Effects of spacing. *Journal of Memory and Language*, *16*, 465.
- Xue, G., Mei, L., Chen, C., Lu, Z., Poldrack, R., & Dong, Q. (2011). Spaced learning enhances subsequent recognition memory by reducing neural repetition suppression. *Journal of Cognitive Neuroscience*, *23*, 1624–1633.
- Yin, J. C. P., Del Vecchio, M., Zhou, H., & Tully, T. (1995). CREB as a memory modulator: Induced expression of a dCREB2 activator isoform enhances long-term memory in drosophila. *Cell*, *81*, 107–115.
- Zechmeister, E. B., & Shaughnessy, J. J. (1980). When you know that you know and when you think that you know but you don't. *Bulletin of the Psychonomic Society*, *15*, 41–44.

- Zhao, X., Wang, C., Liu, Q., Xiao, X., Jiang, T., Chen, C., & Xue, G.** (2015). Neural mechanisms of the spacing effect in episodic memory: A parallel EEG and fMRI study. *Cortex*, **69**, 76–92.
- Zimmerman, J.** (1975). Free recall after self-paced study: A test of the attention explanation of the spacing effect. *American Journal of Psychology*, **88**, 277–291.

Cite this article: 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**, 1103–1139. <https://doi.org/10.1017/S0142716419000158>