Instructional Science (2019) 47:353–371
https://doi.org/10.1007/s11251-019-09483-4

The effects of spacing patterns on incidental L2 vocabulary learning through reading with electronic glosses

Ahmet Çekiç[1] · **Arif Bakla**[1]

Received: 9 November 2018 / Accepted: 13 March 2019 / Published online: 27 March 2019
© Springer Nature B.V. 2019

Abstract

This study investigates the impact of spacing on L2 incidental vocabulary learning. The participants divided into three groups studied the same 20 target words, which occurred nine times in 36 short reading texts in a Moodle course. The first group read four passages and thus saw each target words once per session each week in the nine-week period (nine sessions, fixed spacing group). The second group (three sessions, spaced massing with fixed intervals) and the third group (three sessions, spaced massing with expanding intervals) studied 12 passages each session and saw each target item three times per session. While the second group studied the passages within a seven-week period with a two-week and a three-week interval (three sessions), the third group studied the passages in three consecutive weeks (three sessions). So, all of the three groups saw the same 20 target words a total of nine times with different massing and spacing patterns. The participants were given the Vocabulary Knowledge Scale (the VKS) before and after the treatment and a multiple-choice test immediately after the treatment period. The results indicated that the fixed spacing group outperformed the spaced massing with fixed intervals and spaced massing with expanding intervals.

Keywords Incidental vocabulary learning · Spaced learning · Spaced massing · Hypertext glosses · EFL

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11251-019-09483-4) contains supplementary material, which is available to authorized users.

An earlier version of this paper was delivered as an oral presentation at the 8th International ELT Research Conference, 15–17 May 2014, Çanakkale, Turkey.

✉ Ahmet Çekiç
acekic@cumhuriyet.edu.tr

[1] Foreign Language Education Department, Sivas Cumhuriyet University, Yabancı Diller Egitimi Bolumu, Sivas 58140, Turkey

Introduction

Second language learners have to possess adequate lexical knowledge, which is particularly essential for understanding a wide variety of texts that learners usually come across in their daily and academic lives. Learners usually acquire lexical knowledge through what is commonly referred to as incidental learning, which occurs when the focus of attention is not learning vocabulary items but the comprehension of what is conveyed through written or oral language (Akbulut 2007). In a foreign language class, incidental learning is quite significant because teachers do not have enough time to focus on lexis (Hu et al. 2014); learning happens usually outside the class when learners read and listen to target language texts.

Incidental vocabulary learning, which refers to the acquisition of lexis as a by-product of reading or listening, could promote individualized and authentic learning. Research suggests that it is possible for foreign language learners to learn vocabulary incidentally, especially when they are reading (e.g., Paribakht and Wesche 1999; Webb 2008). It is based on the idea that previously encountered words are committed to long-term memory through exposure in incidental reading.

However, incidental vocabulary learning is not free from problems as several conditions have to be satisfied for such learning to be efficient. First, learners might make incorrect guesses about the meanings of words due to inadequate or ambiguous contextual information, and this may lead to incorrect learning (Hu et al. 2014; Webb 2008). It is rather challenging for foreign language learners, particularly lower-level ones, to acquire the meaning and form of especially newly encountered lexical items by guessing their meanings from the context (Ko 2012). Second, learners should have a working knowledge of a great portion of the words surrounding the target items (Restrepo Ramos 2015), so that they could do the guesswork without much difficulty. Finally, learners need to recycle vocabulary in general. This certainly holds true for incidentally acquired vocabulary items as well. Therefore, learners need to see the target words a number of times for better retention (frequency of exposure).

To deal with these problems, a particular input modification technique is the use of glosses (either paper-based or electronic), which can be used to provide learners with adequate lexical or contextual information. Frequency of occurrence for target vocabulary items in such texts and general frequency of these words in the language itself are two key issues in the acquisition of lexical items. Although a myriad of studies on incidental vocabulary learning have been carried out, it is still unclear how much exposure is needed for a vocabulary item to be learnt (Restrepo Ramos 2015). This is probably because, as Restrepo Ramos notes, there are a number of confounding factors that interfere in incidental learning of vocabulary, such as the characteristics of the target word, features of the context and strategies used by learners for vocabulary guesswork. Therefore, Chen and Truscott (2010) recommend that research should concentrate on such potentially latent variables that could deeply affect learning.

Another significant issue that is as important as the number of occurence is how these occurences are scheduled (Goossens et al. 2012). The way repetitions are scheduled or spread across time is quite significant with respect to the amount of information retained in the long-term memory. The practice of spreading study sessions across time is referred to as "spaced learning" or "distributed learning". Research on spaced learning has found that it is better for learners to learn vocabulary items in multiple sessions, an obvious finding suggested by a number of studies (Bird, 2010 as cited in Carpenter et al. 2012; Goossens et al. 2012; Nakata 2015; Sobel et al. 2011). However, there is a lack of research into

optimal spacing patterns that could lead to more effective and longer lasting learning the way study sessions.

In line with the recommendation of Chen and Truscott (2010) mentioned above, the present study incorporated the issue of spaced learning into incidental vocabulary acquisition studies as it is one of the factors that might have a deep impact on vocabulary learning. To date, very few studies have focused on how the pattern of recycling could impact students' learning and retention of vocabulary through hypertext annotations. Various researchers have highlighted the gaps in spaced learning research: (a) most studies investigating spaced learning have been conducted in controlled laboratory settings by asking learners to memorise lists of words or simple facts (Carpenter et al. 2012), (b) most of the studies were conducted within a short amount of time (ranging from several minutes to a maximum duration of one week) and those exceeding one week are quite limited (Lotfolahi and Salehi 2017). In addition, studies on spaced learning in L2 vocabulary acquisition are quite limited, and existing studies, according to Snoder (2017), have been carried out in the form of paired-associate tasks. Compared with presenting target vocabulary items in context (in a reading text), methods and materials used to investigate spaced learning are poor in ecological validity. Finally, our study is unique in that it incorporates a new spacing pattern into discussion as we blended spacing with massing and named it as "spaced massing".

Review of literature

Vocabulary glosses

Although glosses mostly include definitions, there are a number of techniques used to clarify the meaning of unfamiliar vocabulary, such as L1 equivalents, sample sentences, L2 definitions/synonyms or a combination of these elements or more recently concordance lines (Warschauer and Lee 2017) or glosses for listening texts (Cakmak and Ercetin 2017). Other possible information to be embedded in glosses could be cognates, phonetic transcription, etymological information and so forth. With respect to the medium in which glossary information is provided, Yanguas (2009) divides glosses into three categories: textual, visual and a blend of these. Any text-based information (e.g., L1 translations, L1 or L2 definitions, explanatory notes, synonyms) is included in the textual category, while pictures, audios and videos fall in the second.

To date, a large body of research has been carried out to investigate the impact of hypertext glosses on learning and retention of lexical items in L2 learning contexts. Researchers also focus on type and format of glosses (see Marefat et al. 2016), due to widely accepted benefits of glosses in general, particularly for lower-level learners. In this respect, the focus of attention has shifted to gloss types rather than investigating if they were really useful (Azari 2012; Yoshii 2006). In addition to this, it is easier for material designers to pack more information in online glosses. Such information is there ready to serve the needs of learners without disturbing them. When hypertext is enriched with multimedia, the new tool is referred to as hypermedia, which according to Ariew and Ercetin (2004), facilitate comprehension of reading texts as they include various multimedia elements as text, audio, video, visuals and so forth. Based on an eye-tracking study, Boers et al. (2017) claim that the value that lies in multimedia glosses is that they attract students' attention more than single-mode glosses do. However, multimodal glosses might not always be effective

enough. For example, glosses with L1 equivalents of target words were found to be more effective than those with multimedia elements (Acha 2009).

Other popular gloss-related topics of investigation include the language of glosses. Some studies investigated whether L1 or L2 glosses were better in helping learners acquire vocabulary and retain it, yet results have been inconclusive (Xu 2010) because no significant difference was found in some studies (e.g., Chen 2002; Jacobs et al. 1994; Ko 2012; Yoshii 2006), while in others either L1 or L2 glosses were found to be more effective than the other (Ertürk 2016; Ko 2005; Miyasako 2002). Studies also found that L2 glosses could be better than L1 glosses at more advanced levels (Hu et al. 2014; Ko 2012).

Frequency of exposure

Although it is a well-established finding in the literature that reading is a good source of input and repeated exposure to the target vocabulary items is an essential part of second language vocabulary acquisition (Schmitt 2008), there is no agreement among researchers on the number of exposures essential for learning words incidentally (Restrepo Ramos 2015). Different researchers proposed different numbers for the optimal number of exposures to a vocabulary item, so that it could be committed to long-term memory: seven to ten (Heidari-Shahreza and Tavakoli 2016), between five to 16 encounters when words are presented in appropriate contexts (Nation 1990), and up to 20 encounters (Waring and Takaki 2003). Chen and Truscott (2010), based on findings from Webb (2007), consider seven as the minimum number of repetitions for successful productive vocabulary acquisition. According to Nation (2001), there is no prescribed number of repetitions to learn a word because not only the engagement of the learner but also how the level of learning is assessed are highly influential in prescribing the correct number of repetitions. In short, research on frequency of exposure has brought about inconclusive results. Justifiably, given that frequency alone cannot explain how words are learned as there are several confounding variables that function as predictors of vocabulary acquisition. The time between repetitions is also a significant factor that affects the learning of target vocabulary items (Johnson and Heffernan 2006). The next section elaborates on this issue.

Spaced learning

To date, although a great number of studies have been carried out to investigate the impact of frequency of appearance in a text, they have failed to agree on a common frequency of exposure essential for acquisition. Another issue is the "spacing effect" (Carpenter et al. 2012; Cepeda et al. 2008; Lee and Choe 2014; Nakata 2006, 2008; Smolen et al. 2016), which in plain words refers to the practice of "spreading study sessions in time, instead of cramming them together" (Lotfolahi and Salehi 2017, p. 1). The type of learning in which the spacing effect is taken into account by distributing learning across time is referred to as "spaced learning". Also referred to as "distributed learning" (Lee and Choe 2014; Son and Simon 2012), spaced learning is characterised with varying time periods between repetitions, which are dispersed across the time (Carpenter et al. 2012, p. 370) and lead to better retention than massed learning (Goossens et al. 2012; Smolen et al. 2016). The amount of time left between repetitions in spaced learning could range between seconds/minutes and several weeks, or in some cases it could be months or years. Research on spacing effect suggests that it leads to better retention of information (Cepeda et al. 2006; Son and Simon 2012).

Second Language Acquisition (SLA) research has also suggested that spaced learning works well in language learning, especially in learning grammar (Bird, 2010 as cited in Carpenter et al. 2012) and vocabulary of a foreign language (Goossens et al. 2012; Lotfolahi and Salehi 2017). Similarly, a study by Sobel et al. (2011) found that the group studying words in two sessions with a one-week interval (20.8% retention) outperformed the one who studied the same words in two tutorials with no space between them (7.5% retention) as revealed by the retention test given five weeks after the second session.

Such findings were also supported by various other studies. Kornell (2009) requested undergraduate learners to study 20 word pairs as flashcards (with words and their synonyms). The participants in the spaced learning group studied the word pairs in four sessions (twice in each), while for those in the massed learning group, the 20 words were divided into four groups and each group was studied eight times in a single session, meaning that both groups studied the words eight times through different recycling patterns. The spaced learning group outperformed the massed group by remembering almost twice as many words as the number of words remembered by the latter. Using a within-subjects design, Goossens et al. (2012) compared spaced learning with massed learning of vocabulary in primary school kids (Grade 3, $n=33$). The participants studied the words in four sessions. In the first session, the researcher presented 30 target words in a presentation on the board and tested the students' knowledge of the target words by requesting them to provide their meanings. Then these words were divided into two groups. In the second session, the participants studied the first 15 target words once and they studied the first five of the second set of 15 target words. In the third session the first 15 target words were studied once more and the second set of five words (in the second group of target words) were studied three times, and this was repeated in the last session, with the 15 words in the first group and the last five words of the second group of 15 target words. The participants were tested one week after the final session, and the results indicated that the words studied using the spacing condition were remembered better than those in the massed condition. The researchers note that the participants studied the first set of 15 words first in all three sessions, so this might have led to a stronger spacing effect than they would normally gain (e.g., if two different groups of learners were used).

On the other hand, some other studies found no spacing effect in vocabulary learning. For example, Lee and Choe (2014) studied the impact of spacing on vocabulary learning among beginner Korean EFL learners. They assigned 60 EFL learners to three learning conditions. In the first experiment, the activities were done: (1) in four 10-min sessions with intervals between them (one activity per session), (2) in 20 min sessions with an interval between them (two activities per session), and (3) all for activities in a single 40-min session (with no space between the activities). They replicated the first experiment using a different set of vocabulary items. Results from both experiments revealed that spacing had no effect on learning vocabulary. The researchers attributed this to the low proficiency of the participants. In another study, Snoder (2017) worked with 45 Swedish-L1 learners of English, who studied 28 target collocations through what he/she calls intensive learning (repetitions on day 1, 2, 4) or expanded learning (repetitions on day 1, 7 and 16). The results indicated that spacing did not have a significant effect on learning gains.

Given that the benefit of spacing is almost obvious, we could now turn our attention to the way repetitions or revisions are scheduled, which according to Nakata (2006), might be highly influential in the amount of vocabulary learned. A significant pair of spacing-related terminology are "absolute spacing" and "relative spacing." While the former is used to mean the total amount of spacing used between all the repetitions, the latter refers to how study sessions are scheduled across time (Nakata 2013). In this sense, spacing effect seems

to influence learning as much as frequency of exposure does. This is because when a small list of vocabulary items is repeated 10 times in a single session or the same list is repeated twice in five different sessions distributed across time, based on research findings mentioned above, one would accept that the second method will be more successful.

Another issue that research focussed on was the length of space left between study sessions. The gaps separating study sessions could be fixed (equal amount of time between repetitions, such as one-day gaps between each study session) or expanding (gradually increasing amounts of time between repetitions, such as 30 min, one day, two weeks) (Carpenter et al. 2012). In another study, Khoii and Abed (2017) investigated the impact of no space, fixed and expanding spacing on receptive and productive vocabulary knowledge. They requested 63 female Iranian EFL learners to study 20 word pairs (words and their synonyms). They gave the students a vocabulary knowledge scale to test the impact of the three spacing conditions and found that the equal spacing group outperformed the others in both receptive and productive lexical knowledge. Carpenter et al. (2012) further note that regardless of type of spacing (i.e., fixed or expanding), spacing is superior to massing (or cramming) of the learning material. Probably due to inconclusive research results on fixed and expanding spacing, they note that what is significant is to leave some gaps between study sessions rather than to pay a lot of attention to if the spaces are fixed or expanding.

Research questions

Based on the research gaps in the literature mentioned above, the present study sought to find answers to the following research questions:

(RQ1) How do different spacing patterns affect the receptive and productive vocabulary gains?
(RQ2) What is the impact of the length of the total time allocated for all the learning sessions on receptive and productive vocabulary gains?

In line with these research questions, we hypothesize that the fixed spacing group which studies the words in a nine-week period will outperform the other experimental groups as it involves consistent spacing without cramming. Furthermore, we also think that the spaced massing group with expanding intervals will perform better than the spaced massing group with fixed intervals as the former group was exposed to the target words for a longer period of time (seven weeks as compared to three weeks).

Materials and methods

The present study investigated the impact of spacing patterns on learning and retention of English vocabulary by intermediate learners. The study adopted an experimental research design without a control group. The study group was composed of 189 intermediate Turkish-L1 learners of English, and the researchers used random cluster sampling to divide the participants into three experimental groups. The participants were exposed to 20 lexical items in 36 short passages, and while the timing of exposure varied across the three groups, the number of exposure was kept constant. The first experimental group, which can be called as fixed spacing condition, read four passages a session, so they were exposed to target words once a week for a nine-week period (henceforth referred to as Pattern 9).

The second experimental group read 12 passages in each session within seven weeks with two-week intervals between sessions (henceforth referred to as Pattern 7), which can be called "spaced massing" with expanding intervals. The participants in the third experimental group studied the target words three times in each session, which signals massing in learning condition. There were also one-week spacing between each session; therefore, the researchers opt to use a term indicating the blended nature of the learning condition: "spaced massing" with fixed intervals (henceforth referred to as Pattern 3). So the participants in all groups were exposed to each target word three times in each session (Fig. 1). Pure massed learning, which earlier studies consistently found to be less conducive to incidental learning than spacing (e.g., Goossens et al. 2012; Kornell 2009; Smolen et al. 2016; Sobel et al. 2011), was not used in this study. The participants were given only the vocabulary knowledge test (VKS) because it was considered to be a better tool for testing vocabulary items which were mostly unknown to the participants (Online Appendix A).

Participants

A total of 189 Turkish-L1 learners of English (independent users, Common European Framework of Reference for Languages -CEFR- levels) registered at an intermediate reading class at an English preparatory program were invited to take part in the study. However, not all of the participants agreed to take part in the study and not all the volunteers were able to attend all of the sessions and tests. Some students were eliminated because they failed to take any one of the tests or missed a session. Thus, we ended up with a total of 77 students who met our data inclusion criteria, as it was the case in some other studies (e.g., Goossens et al. 2012; Snoder 2017). The data from some participants were excluded,

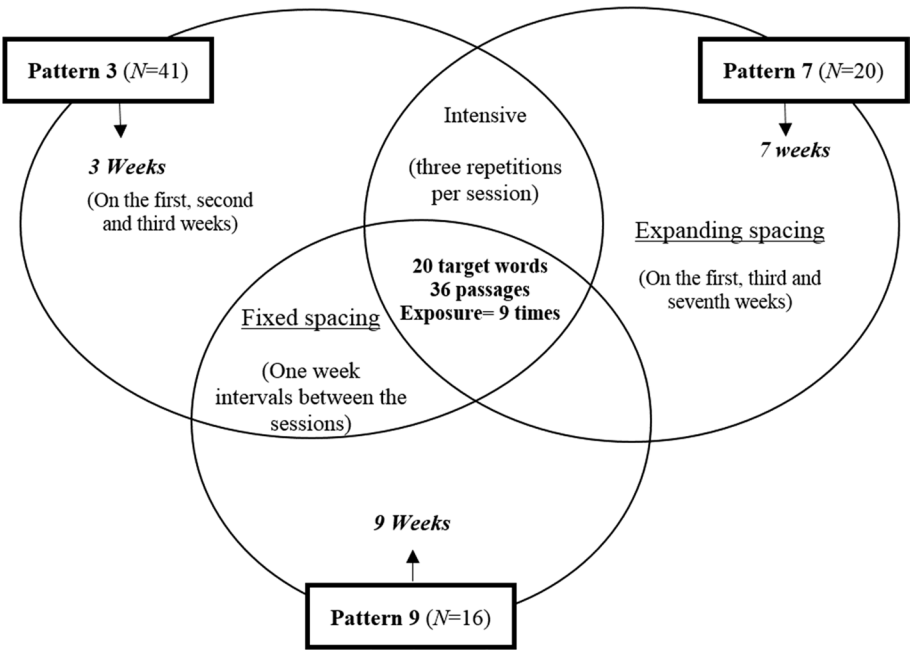


Fig. 1 The patterns of exposure to the target vocabulary items in the three experimental groups

Table 1 Text analysis statistics for the reading materials used in the study

Text statistics	Corresponding value
Word count	3841
Average word count (per passage)	106.70
Unique word count	1200
Total sentence count	193
Average sentence count (per passage)	5.36
Total passage count	36

Table 2 Readability statistics for the reading materials used in the study

Readability formula	Grade
Flesch-Kincaid grade level	11.3
Gunning Fog index	14.2
Coleman-Liau index	11.2
Flesch reading ease	49.8
CEFR level	B2
IELTS level	5–6

indicating a drawback of conducting a long-term study on spacing effect. The participants read the texts on the computer in a laboratory.

Hypertext reading materials

The authors used a total of 36 short passages (with an average word count of 106.70). To keep the frequency of exposure constant, the authors wrote 32 of these passages as in some other studies (Chen and Truscott 2010; Heidari-Shahreza and Tavakoli 2016; Rott 1999). Thus, each one of the 20 target words appeared nine times in the passages (Table 1).

The texts were of average difficulty, suggesting that they were readable for independent users (CEFR level) (Table 2). The texts were written in formal language and covered a variety of topics ranging from business life to history and from technology to health.

The texts were presented in nine scorm packages, produced using a commercial authoring tool (Softchalk v9). The scorm packages included target vocabulary items annotated with audio pronunciation of the words and their audio and written L2 definitions. The audios were recorded by the researchers themselves. Tooltip-type glosses (Lee and Lee 2015) were used. That is, the participants were able to see the definition of the word when they pointed to the words emphasised using a different colour and boldfacing. They were able to hear the related definition or synonym when they clicked on the speaker icon next to each vocabulary item written in bold. All the vocabulary items were presented in a context, and each passage was accompanied by two multiple-choice reading comprehension questions. Upon answering these questions, the participants were given immediate feedback,

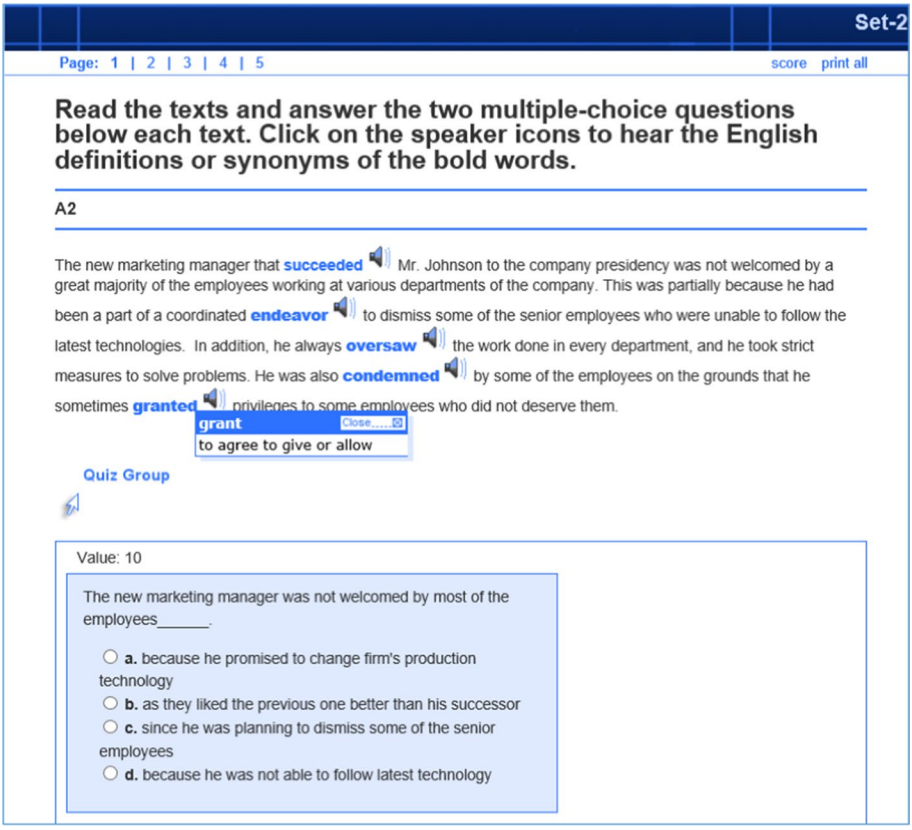


Fig. 2 A sample reading passage with audio and written hypertext glosses and a multiple-choice comprehension question is shown below. *Note* The screen is able to show only one of the two multiple-choice items

and the scorm packages in the learning management system recorded the answers of each student both in the scorm package and in the gradebook for the course (Fig. 2).

The target vocabulary items were chosen according to proficiency level of learners based on the researchers' judgements, who regularly teach almost half of the participants. In addition, a vocabulary profile software was used to analyse the frequency of the potential target words, which were profiled using Cobb (n.d.) (BNC-COCA, provided in the software). Both lower and higher frequency words were chosen because this practice could help boost the ecological validity of the learning tasks (Webb et al. 2013) (Table 3). This, along with selecting real words instead of non-words or relatively low-frequency words, helped establish a balance between control and ecological validity. Of the higher frequency words (K-2 and K-3 word families), some were used due to their less frequently known meanings (e.g., the verb "contract" was used to mean "to catch an illness" or the verb "succeed" was used to mean "take over a throne, office, or other position from somebody"). A note about this was added to the Vocabulary Knowledge Scale to warn the participants to disregard the commonly known meanings of two target items (See Online Appendix A, Item 5 and 18).

Table 3 Frequency analysis of the target words selected

Frequency level	Families (%)	Words
K-2 words	3 (15.00)	Contract, grand, shed
K-3 words	4 (20.00)	Cite, condemn, succeed, pursue
K-4 words	2 (10.00)	Endeavour, oversee
K-5 words	2 (10.00)	Jeopardy, relish
K-6 words	5 (25.00)	Afflict, drawback, evade, proliferate, tout
K-7 words	2 (10.00)	Compulsion, sedative
K-10 words	1 (05.00)	Inkling
K-11 words	1 (05.00)	Malady

Data collection tools

The researchers collected data by using two quantitative data collection tools: (a) Vocabulary Knowledge Scale (the VKS), adapted from Wesche and Paribakht (1996) and (b) a 20-item multiple-choice test for the target lexical items (developed and validated by the researchers themselves) (Appendix B). Although the receptive vocabulary test was not sensitive to partial learning of the target items, the VKS is a more sensitive tool that could at least partially test various dimensions of vocabulary knowledge, such as form, meaning, part of speech and productive knowledge.

The researchers developed a multiple-choice test, and it was piloted with upper-intermediate and advanced students. They made sure that the given context for each target item in the multiple-choice test was clear, and no unfamiliar words were used. Moreover, they asked a native speaker to do the test and suggest revisions for each item. The native speaker answered all of the questions correctly and suggested revisions to make the context clearer. The revised version was used as the post-test.

Data analysis

The participants' scores for the VKS were calculated based on the scheme, suggested by the originators of the scale (Wesche and Paribakht 1996) (See Fig. 3) and were compared using MANOVA. The reading comprehension scores were only used to identify if the students read the passages or not. Normality tests for all the datasets (pre-test VKS, post-test VKS scores, receptive vocabulary scores) were carried out. The alpha level was considered to be 0.05 in all statistical analyses, except for the cases in which the Bonferoni correction was used (e.g., in MANOVA or follow-up Mann–Whitney *U* tests in the Kruskall Wallis test) (The significance level divided by 3: 0.05/3 = 0.0167).

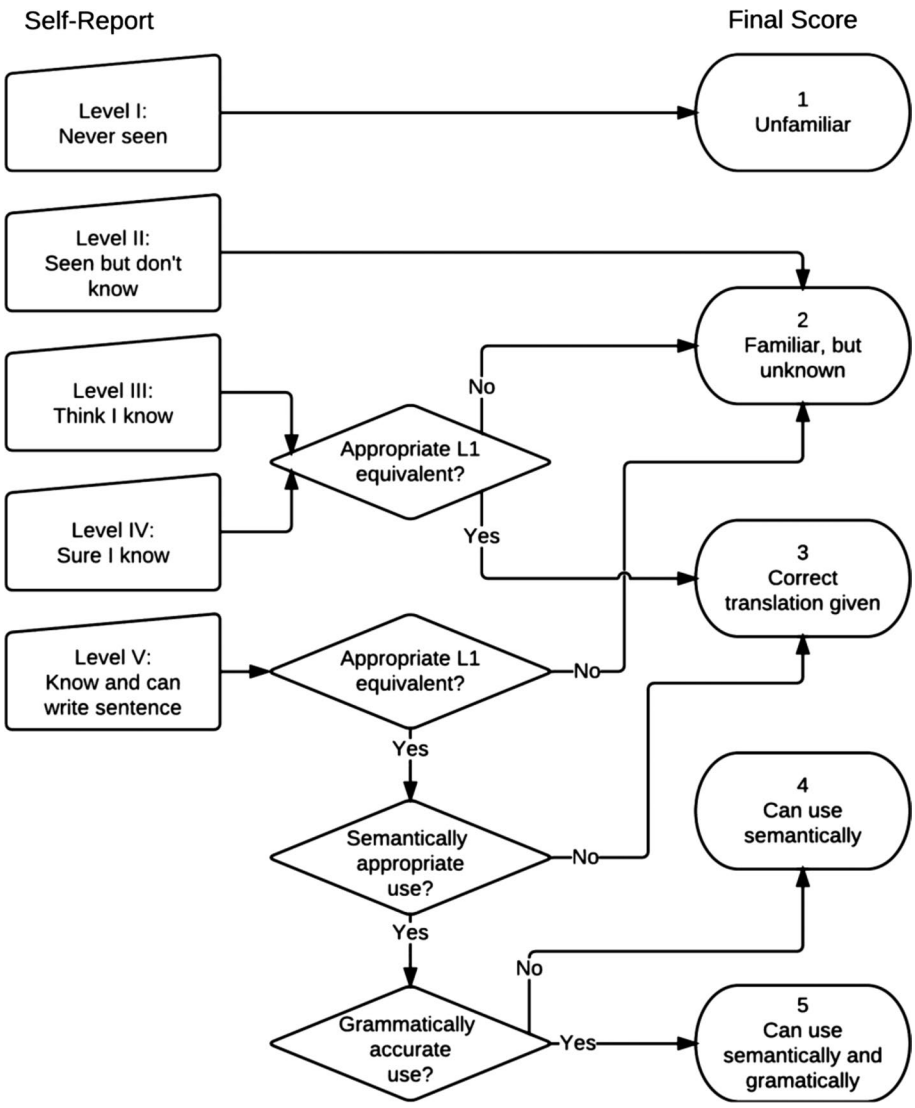


Fig. 3 The procedure followed while calculating the VKS scores (Adapted from Wesche and Paribakht 1996)

Results

Results from the productive vocabulary test (VKS)

The participants improved their productive vocabulary knowledge from the pretest to posttest (VKS) (Table 4). To examine if this improvement was significant, a within-subjects repeated measures ANOVA was conducted to compare the effect of incidental vocabulary learning on the participants' test scores across the two testing times.

Table 4 Statistics about the study group's scores across the two testing points

Test	Group	N	Mean	SD
Pretest (VKS scores)	Pattern 3	41	38.05	8.10
	Pattern 7	20	37.30	4.96
	Pattern 9	16	37.44	9.82
	Global	77	37.73	7.74
Posttest (VKS scores)	Pattern 3	41	66.78	17.53
	Pattern 7	20	52.75	14.58
	Pattern 9	16	71.06	18.01
	Global	77	64.03	18.08

First, normality was checked. Skewness and kurtosis scores were found to be within + 1.96 and − 1.96 for the dependent variables, and the related data were checked for outliers because repeated measures ANOVA is known to be sensitive to lack of normality which is caused by outliers, yet there were no outliers in the data. In addition, we tested homogeneity using Box's *M* test and Levene's test, and Box's *M* test was not significant, ($M = 10.131$, $F = 1.608$, $p = 0.140$). However, although Levene's test was not significant for the posttest, it was significant for the pretest, meaning that the assumption of homogeneity of error variances was not met for the former. Therefore, the partial violation of equality of error variances is acknowledged as a limitation.

As it is a robust test against the violation of assumptions of ANOVA, Pillai's Trace test was used, and it was found to be significant, 0.717, $F(1, 74) = 187.75$, $p = 0.001$, multivariate $\eta_p^2 = 0.717$, power = 1.000. This showed that the mean scores differed across the pre-test and post-test VKS scores at statistically significant levels. The effect size of the improvement across the two testing times was quite large. This indicated that nearly 72% of multivariate variance in the dependent variables is associated with the time factor.

Pairwise comparisons were carried out using the Bonferroni correction. They revealed that the instruction produced a significant improvement in the participants' productive vocabulary test scores (VKS) from the pretest ($M = 37.73$, $SD = 7.74$) to the posttest ($M = 64.03$, $SD = 18.08$), which was statistically significant ($p = 0.001 < 0.05$). Globally

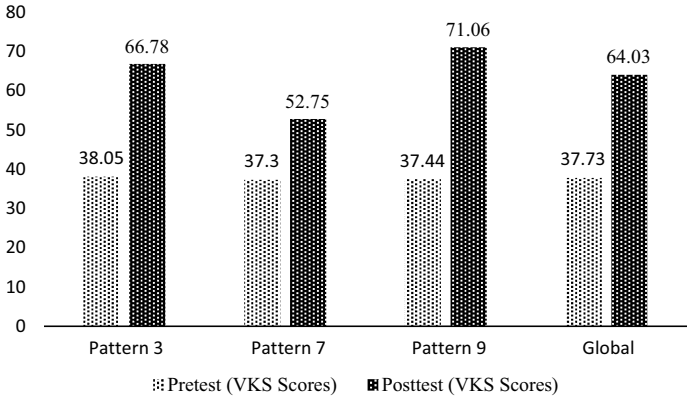


Fig. 4 Within-subject comparisons of the VKS scores across the pretest and posttest

considered, the instructional practice in the study helped boost the participants' scores at statistically significant levels (Fig. 4).

To examine the differences between the pretest and posttest between the experimental groups in the repeated measures ANOVA, the interaction effect (time*group) was tested. Pillai's Trace test was conducted, and it was significant, 0.165, $F(2, 74) = 7.34$, $p = 0.001$, multivariate $\eta_p^2 = 0.165$, power = 0.929. This finding indicates that the change between the groups is significantly different and the effect size was medium. The results of the repeated measures ANOVA showed that there was a change between the two testing times. However, the results did not show if the mean scores of the groups differed across the pretest and posttest. A one-way MANOVA was conducted to examine this.

The sufficiency of the sample size was met as one of the assumptions given that the number of smallest group was larger than the number of levels in the independent variable multiplied by the number of dependent variables ($3 \times 2 = 6$) (the smallest group, pattern 9, was composed of 16 participants). The pre-test VKS and post-test VKS scores were independent from each other. The data were initially checked for multivariate normality. There were no outliers in the data. The maximum Mahalanobis distance was within the allowable limit ($D^2 = 9.65$) because it could be up to 13.82 for a MANOVA with two dependent variables. The scatter/dot diagrams indicated that the linear relationship assumption was met. Moreover, Box's M test of equality of covariance matrices was not significant, ($M = 10.131$, $F = 1.608$, $p = 0.140$). There was a moderate significant correlation between the pre-test VKS and post-test VKS scores, $r\ (77) = 0.40$, $p = 0.001$, which meant that multicollinearity between these variables did not exist. It should be noted that Mauchly's W test was not conducted as there were two dependent variables. The results of the multivariate tests were used.

Because MANOVA is a robust test and there is no non-parametric equivalent for it, it was conducted in spite of small violations of its assumptions. Moreover, Pillai's Trace test (again a robust test against the violation of assumptions) was used in MANOVA, which provided a significant multivariate effect for the patterns of spacing (Pattern 3, 7 and 9) on the test scores, Pillai's trace = 0.168, $F(2, 74) = 3.397$, $p = 0.011$, $\eta_p^2 = 0.84$, power = 0.842 (a large effect size). This meant that the mean VKS scores significantly differed across the pretest and the posttest. The effect size of the change across the two testing times was quite large.

As the overall test was significant, the univariate main effects were examined. On the one hand, no significant differences were found for the pretest, $F(2, 74) = 0.075$, $p = 0.928$, $\eta_p^2 = 0.002$, power = 0.061. This indicated that their mean scores were approximately equal before the instruction. On the other hand, univariate main effects were significant for the posttest $F(2, 74) = 6.364$, $p = 0.003$, $\eta_p^2 = 0.147$, power = 0.889. The effect size was very small in the former, while it was moderate in the latter (Table 5).

Post hoc comparisons of the experimental groups in the posttest were conducted using Scheffe, due to unequal group sizes. A significant difference was found in the post-test VKS scores, between Pattern 3 ($M = 66.78$, $SD = 17.53$) and Pattern 7 ($M = 52.75$, $SD = 14.58$),

Table 5 Results of the multivariate analysis of variance

Test	F	p	η_p^2	Power
VKS pretest	0.075	*0.928*	0.002	0.061
VKS posttest	6.364	*0.003**	0.147	0.889

*$P < .05$

$p=0.013$ and between Pattern 7 ($M=52.75$, $SD=14.58$) and Pattern 9 ($M=71.06$, $SD=18.01$), $p=0.008$, but there was not a statistically significant difference between Pattern 3 ($M=66.78$, $SD=17.53$) and Pattern 9 ($M=71.06$, $SD=18.01$), $p=0.693$. These results from the repeated measures ANOVA and MANOVA altogether indicate that Pattern 9 and Pattern 3 made (almost equal) improvements in their scores and their scores were better than Pattern 7 at statistically significant levels.

Results from the receptive vocabulary test (multiple-choice)

To compare the post-test scores of the groups a Kruskal–Wallis test was used as two assumptions of ANOVA for the receptive post-test data were violated (normal distribution and Levene's test of equality of variances). Before carrying out the test, the assumptions were checked and it was seen that all of them were met (independent data points and random selection). The test indicated that there was a statistically significant difference between the three spacing patterns (Pattern 3, 7, and 9) $H(2)=16.441$, $p=0.001$, with a mean rank of 34,46 for Pattern 3, 32,28 for Pattern 7 and 59,03 for Pattern 9.

Separate Mann–Whitney U tests were carried out for pairwise comparisons (with a Bonferoni correction, $0.05/3=0.0166$.). The tests indicated that there was not a significant difference between Pattern 3 ($Mdn=12$) and Pattern 7 ($Mdn=13$), $U=402.500$, $p=0.908$, $r=0.0{,}02$ (a very small effect size), while there was a statistically significant difference, not only between Pattern 3 ($Mdn=12$) and Pattern 9 ($Mdn=18$), $U=134.500$, $p=0.001$, $r=0.46$ (a moderate effect size), but also between Pattern 7 ($Mdn=13$) and Pattern 9 ($Mdn=18$), $U=33.000$, $p=0.001$, $r=0.68$ (a large effect size).

Discussion

This study investigated the impact of spacing patterns on receptive and productive vocabulary gains for 20 target items. A notable difference between earlier studies (e.g., Goossens et al. 2012; Lotfolahi and Salehi 2017; Smolen et al. 2016; Sobel et al. 2011; Son and Simon 2012) and the present one was that no massed learning was used in the current study. Following from the study of Snoder (2017), our study provided more comparative data on intensive, expanding and long-term fixed spacing. The data from the productive vocabulary knowledge test (VKS) indicated that the productive vocabulary knowledge of the whole study group increased at a significant level from the pretest to the posttest. This indicated that nine repetitions proved highly effective in boosting the participants' vocabulary knowledge. With respect to both productive and receptive vocabulary gains, it was Pattern 9 that provided the best results. In line with the three research questions formulated earlier, a more detail discussion of the results is presented below.

(Q1) How do different spacing patterns affect the receptive and productive vocabulary gains?

A particular finding of the present study is that fixed spaces between study sessions seems to be better than expanding spacing; a moderate effect size was calculated for this. We should note that Pattern 7 (spaced massing with expanding intervals) studied the words in seven weeks in total. Particular difference between Pattern 7 and Patterns 3 and 9 was that the former studied the words based on a schedule of expanding spacing. Pattern 3

(spaced massing with fixed intervals) and Pattern 9 were similar in this respect because they studied the words in sessions with one week intervals between each, although the number of repetitions differed (three vs. one repetitions per session, respectively).

With respect to receptive vocabulary gains, the participants in Pattern 9 outperformed both those in Pattern 3 and in Pattern 7, lending support to some studies that found a positive effect of spacing. On the other hand, there was no statistically significant difference between the scores of the participants in Pattern 3 and those of the participants in Pattern 7. As for the productive vocabulary gains, the participants who studied the words with one-week intervals between study sessions (Pattern 3 and 9) outperformed those who studied words in an expanding schedule of repetition (Pattern 7). In other words, fixed spacing proved better than expanding spacing because the scores of the participants in Pattern 9 and Pattern 3 were higher than those of Pattern 7 at statistically significant levels, but no significant differences were found between Pattern 9 and Pattern 3. This might suggest that vocabulary revision at regular intervals leads to higher gains in productive vocabulary knowledge. In his/her study, Snoder (2017) attributed the lack of statistically significant differences between intensive and spaced learning conditions to the fact that the former was a pure massed condition. Although Pattern 3 and Pattern 7 were not pure massed learning conditions, there was a difference between the long-term fixed learning condition (Pattern 9) and the other two. This could suggest that there could be a notable difference not only between pure massed learning and spaced learning, but also between intensive and spaced learning, and the best practice seems to be long-term spaced learning with fixed intervals between each session.

On the other hand, longer space between the sessions in Pattern 7 (two weeks between sessions one and two and three weeks between sessions two and three) might have had a negative effect on their learning. This seems highly likely as the major difference between Pattern 7 and Patterns 3 and 9 was that Pattern 7 was characterised with expanding spacing, while the other two patterns were characterised with fixed spacing between the learning sessions. The longest spacing between learning sessions (expanding) was in Pattern 7, and the participants in this group produced the poorest scores. The results of the present study suggest that too much spacing between learning sessions might not be as effective as leaving shorter and regular spaces between them. This was not the case in the other two learning conditions as they repeated the target words regularly. Poorer performance in Pattern 7 seems to support Carpenter et al.'s (2012) warning that a lot of time between learning sessions might have a negative impact on learning vocabulary. It should be noted that the participants in all groups were given the posttests one week after the final session, which rules out the impact of testing time.

(Q2) What is the impact of the length of the total time allocated for all the learning sessions on receptive and productive vocabulary gains?

The total amount of time allocated for the whole task varied across the three groups; as their respective names suggest, the participants in Pattern 9 read the 36 passages and repeated the target words nine times, in nine weeks. Pattern 7 did the same thing in seven weeks, while Pattern 3 did it in three weeks. We hypothesised, earlier in the study, that Pattern 9 would outperform the other two groups in both receptive and productive tests, and we expected there might also be a difference between Pattern 7 and Pattern 3 in favour of the former. The results from the receptive vocabulary test revealed that Pattern 9 was better than both Pattern 3 and Pattern 7, suggesting that long-term fixed spacing

is more effective than both spaced massing with fixed intervals and spaced massing with expanding intervals. The large effect size for the comparison between Pattern 7 and Pattern 9 indicates that massed sessions with expanding space with them is not effective in comparison with long-term fixed spacing, but spaced massing in a shorter time could a viable option for receptive vocabulary acquisition although the best method seems to be long-term fixed spacing.

For the productive vocabulary knowledge test, Pattern 7 was the least successful group. With respect to the total amount of time allocated for the whole task, Pattern 7 was close to Pattern 9. Although Pattern 3 was intensive, it provided better results than Pattern 7, which lasted longer. There was no statistically significant difference between Pattern 3 and Pattern 9, which was an unexpected result. Spacing between massed sessions could have led to such a result. As noted by Son and Simon (2012), students tend to prefer massed learning to spaced learning, meaning that they might have remained motivated during the three weeks in Pattern 3. When the results from the receptive and productive vocabulary test considered globally, too much space between massed sessions is not effective (as in Pattern 7).

When learning sessions are dispersed across time (with equal spacing between them), both productive and receptive vocabulary gains increase. In other words, regular repetitions in a longer period prove to be better than shorter (and probably irregular) ones. In short, the results of the study suggested that, the number of repetitions being equal, completing the nine repetitions in nine weeks proved better than the other two learning conditions (in seven weeks and three weeks) for receptive vocabulary knowledge, and better than Pattern 7 for productive vocabulary learning. However, contrary to our expectations when the two massing groups, Pattern 7 and Pattern 3, are compared, it was seen that Pattern 3 outperformed Pattern 7, which can be attributed to the irregular and expanding spaces between sessions. Globally considered, both the number of repetitions and how these repetitions are spaced are two important aspects that need special attention in vocabulary instruction.

Conclusion

As found in the present study and suggested by previous research, spacing is more beneficial than massing in vocabulary learning. However, as noted by Son and Simon (2012), students prefer massed learning to spaced learning. Therefore, raising awareness of the value of spaced learning is quite significant. The results of the present study also suggest that spaced massing condition could be another option for vocabulary learning although more research is needed on this issue. Moreover, language educators should carefully monitor the impact of expanding the space between learning sessions for learning vocabulary. Poorer performance of the participants in Pattern 7 supports this. Distributing the learning sessions in a longer time was found to be effective, yet the element of regular spacing between sessions seems to be significant. Therefore, learners are recommended to study vocabulary using long-term fixed spacing (the most successful group in the present study). In the present study, the students learned the words incidentally, so it is not certain if the results could apply to deliberate learning of vocabulary items. However, our results seem to have the potential to inform how frequently target vocabulary items could appear in course books. In this respect, the results globally suggested that not only the frequency of words but also the spacing between the occurrences of the target words should be considered in learning materials. They also indicated that spacing and the total amount of time allocated

for all the repetitions work in tandem. Therefore, one has to consider both constructs when recycling vocabulary is the issue.

Limitations

The researchers could have organized Patterns 3 and 7 as three different sessions for each week so that students could read four passages each session rather than reading 12 passages at once. Another limitation of the study was that, due to reasons of time, no late posttest was given, so long-term impacts of the spacing effect across the three groups were not tested. Unlike some researchers did (e.g., Webb, 2008), we did not test other essential aspects of lexical knowledge other than meaning. During the statistical analyses, despite the violation of equality of error variances, the repeated measures ANOVA was carried out. Finally, the researchers did not calculate an interrater reliability score for the VKS scores as only one person rated them.

Further research

As noted by Goossens et al. (2012), prospective studies could investigate if spacing effect could help reduce the number of repetitions essential for the acquisition of vocabulary. A particular area of further research could be to incorporate not only expanding spacing but also irregular spacing (revising vocabulary items at intervals with unequal spaces between them). In this respect, the comparison of regularly and irregularly spaced learning sessions could be another topic of further investigation. Furthermore, more research could be carried out to compare massed, intensive and expanding learning on vocabulary learning. The present study compared spaced massing with fixed intervals, spaced massing with expanding intervals and long-term spacing with fixed intervals between them. Prospective studies could compare completely massed learning with intensive learning. As noted in the limitations, usually due to absences, the number of participants in Patterns 9 and 7, and partially in Pattern 3, decreased significantly towards the end of the study. Therefore, there was a small number of participants in two of the patterns. Future studies could replicate the present study with a larger number of participants.

References

Acha, J. (2009). The effectiveness of multimedia programmes in children's vocabulary learning. *British Journal of Educational Technology, 40,* 23–31. https://doi.org/10.1111/j.1467-8535.2007.00800.x.

Akbulut, Y. (2007). Effects of multimedia annotations on incidental vocabulary learning and reading comprehension of advanced learners of english as a foreign language. *Instructional Science, 35,* 499–517. https://doi.org/10.1007/s11251-007-9016-7.

Ariew, R., & Ercetin, G. (2004). Exploring the potential of hypermedia annotations for second language reading. *Computer Assisted Language Learning, 17*(2), 237–259. https://doi.org/10.1080/0958822042000334253.

Azari, F. (2012). Review of effects of textual glosses on incidental vocabulary learning. *International Journal of Innovative Ideas, 12*(2), 13–24.

Boers, F., Warren, P., Grimshaw, G., & Siyanova-Chanturia, A. (2017). On the benefits of multimodal annotations for vocabulary uptake from reading. *Computer Assisted Language Learning, 30*(7), 709–725. https://doi.org/10.1080/09588221.2017.1356335.

Cakmak, F., & Ercetin, G. (2017). Effects of gloss type on text recall and incidental vocabulary learning in mobile-assisted L2 listening. *ReCALL, 30*(1), 24–47. https://doi.org/10.1017/S0958344017000155.

Carpenter, S. K., Cepeda, N. J., Rohrer, D., Kang, S. H., & Pashler, H. (2012). Using spacing to enhance diverse forms of learning: Review of recent research and implications for instruction. *Educational Psychology Review, 24*(3), 369–378. https://doi.org/10.1007/s10648-012-9205-z.

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. https://doi.org/10.1037/0033-2909.132.3.354.

Cepeda, N. J., Vul, E., Rohrer, D., Wixted, J. T., & Pashler, H. (2008). Spacing effects in learning: A temporal ridgeline of optimal retention. *Psychological Science, 19*(11), 1095–1102. https://doi.org/10.1111/j.1467-9280.2008.02209.x.

Chen, H. (2002). *Investigating the effects of L1 and L2 glosses on foreign language reading comprehension and vocabulary retention.* Paper presented at the annual meeting of the Computer-Assisted Language Instruction Consortium, Davis, CA.

Chen, C., & Truscott, J. (2010). The effects of repetition and L1 lexicalization on ıncidental vocabulary acquisition. *Applied Linguistics, 31*(5), 693–713. https://doi.org/10.1093/applin/amq031.

Cobb, T. (n.d.) *Compleat Web VPv.2 [Computer software]*. https://www.lextutor.ca/vp/comp/

Ertürk, Z. Ü. (2016). The effect of glossing on EFL learners' ıncidental vocabulary learning in reading. *Procedia: Social and Behavioral Sciences, 232,* 373–381. https://doi.org/10.1016/j.sbspro.2016.10.052.

Goossens, N. A. M. C., Camp, G., Verkoeijen, P. P. J. L., Tabbers, H. K., & Zwaan, R. A. (2012). Spreading the words: A spacing effect in vocabulary learning. *Journal of Cognitive Psychology, 24*(8), 965–971. https://doi.org/10.1080/20445911.2012.722617.

Heidari-Shahreza, M. A., & Tavakoli, M. (2016). The effects of repetition and L1 lexicalization on incidental vocabulary acquisition by Iranian EFL learners. *Language Learning Journal, 44*(1), 17–32. https://doi.org/10.1080/09571736.2012.708051.

Hu, S. M., Vongpumivitch, V., Chang, J. S., & Liou, H. C. (2014). The effects of L1 and L2 e-glosses on incidental vocabulary learning of junior high-school English students. *ReCALL, 26*(1), 80–99. https://doi.org/10.1017/S0958344013000244.

Jacobs, G. M., Dufon, P., & Hong, F. C. (1994). L1 and L2 vocabulary glosses in L2 reading passages: Their effectiveness for increasing comprehension and vocabulary knowledge. *Journal of Research in Reading, 17*(1), 19–28. https://doi.org/10.1111/j.1467-9817.1994.tb00049.x.

Johnson, A., & Heffernan, N. (2006). The short readings project: A CALL reading activity utilizing vocabulary recycling. *Computer Assisted Language Learning, 19*(1), 63–77. https://doi.org/10.1080/09588220600804046.

Khoii, R., & Abed, K. F. (2017). Effects of equal spacing, expanding spacing, and massed condition on EFL learners' receptive and productive vocabulary retrieval. In Pixel, Proceedings of ICT for language learning (10th Ed.) ICT for language learning. (pp. 500–504) 9–10 November, 2017, Florence, Italy.

Ko, M. H. (2005). Glosses, comprehension, and strategy use. *Reading in a Foreign Language, 17*(2), 125–143.

Ko, M. H. (2012). Glossing and second language vocabulary learning. *TESOL Quarterly, 46*(1), 56–79. https://doi.org/10.1002/tesq.3.

Kornell, N. (2009). Optimising learning using flashcards: Spacing is more effective than cramming. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition, 23*(9), 1297–1317. https://doi.org/10.1002/acp.1537.

Lee, E., & Choe, M.-H. (2014). The effect of spaced repetitions on Korean elementary students' L2 English vocabulary learning. *Studies in English Education, 19*(1), 55–75.

Lee, H., & Lee, J. H. (2015). The effects of electronic glossing types on foreign language vocabulary learning: Different types of format and glossary information. *Asia-Pacific Education Researcher, 24*(4), 591–601. https://doi.org/10.1007/s40299-014-0204-3.

Lotfolahi, A. R., & Salehi, H. (2017). Spacing effects in vocabulary learning: Young EFL learners in focus. *Cogent Education, 4*(1), 1–10. https://doi.org/10.1080/2331186X.2017.1287391.

Marefat, H., Rezaee, A. A., & Naserieh, F. (2016). Effect of computerized gloss presentation format on reading comprehension: A cognitive load perspective. *Journal of Information Technology Education Research, 15,* 479–501. https://doi.org/10.28945/3568.

Miyasako, N. (2002). Does text-glossing have any effects on incidental vocabulary learning through reading for Japanese senior high school students? *Language Education & Technology, 39,* 1–20.

Nakata, T. (2006). Implementing optimal spaced learning for English vocabulary learning: Towards improvement of the low-first method derived from the reactivation theory. *The JALT CALL Journal, 2*(2), 3–18.

Nakata, T. (2008). English vocabulary learning with word lists, word cards and computers: Implications from cognitive psychology research for optimal spaced learning. *ReCALL, 20*(1), 3–20. https://doi.org/10.1017/S0958344008000219.

Nakata, T. (2013). *Optimising second language vocabulary learning from flashcards.* (Unpublished doctoral dissertation). Victoria University of Wellington, New Zealand.
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*(4), 677–711. https://doi.org/10.1017/S0272263114000825.
Nation, I. S. P. (1990). *Teaching and learning vocabulary*. New York, NY: Newbury House.
Paribakht, T. S., & Wesche, M. (1999). Reading and "incidental" L2 vocabulary acquisition. *Studies in Second Language Acquisition, 21,* 195–224. https://doi.org/10.1017/S027226319900203X.
Restrepo Ramos, F. D. (2015). Incidental vocabulary learning in second language acquisition: A literature review (Aprendizaje incidental de vocabulario en la adquisición de una segunda lengua: una revisión de literatura). *PROFILE: Issues in Teachers' Professional Development, 17*(1), 157–166. https://doi.org/10.15446/profile.v17n1.43957.
Rott, S. (1999). The effect of exposure frequency on ıntermediate language learners' incidental vocabulary acquisition and retention through reading. *Studies in Second Language Acquisition, 21*(4), 589–619. https://doi.org/10.1017/S0272263199004039.
Schmitt, N. (2008). Instructed second language vocabulary learning. *Language Teaching Research, 12*(3), 329–363. https://doi.org/10.1177/1362168808089921.
Smolen, P., Zhang, Y., & Byrne, J. H. (2016). The right time to learn: mechanisms and optimization of spaced learning. *Nature Reviews Neuroscience, 17*(2), 77–88. https://doi.org/10.1038/nrn.2015.18.
Snoder, P. (2017). Improving English learners' productive collocation knowledge: The effects of ınvolvement load, spacing, and ıntentionality. *TESL Canada Journal, 34*(3), 140–164. https://doi.org/10.18806/tesl.v34i3.1277.
Sobel, H. S., Cepeda, N. J., & Kapler, I. V. (2011). Spacing effects in real-world classroom vocabulary learning. *Applied Cognitive Psychology, 25*(5), 763–767. https://doi.org/10.1002/acp.1747.
Son, L. K., & Simon, D. A. (2012). Distributed learning: Data, metacognition, and educational implications. *Educational Psychology Review, 24*(3), 379–399. https://doi.org/10.1007/s10648-012-9206-y.
Waring, R., & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign language, 15*(2), 130–163.
Warschauer, M., & Lee, J. H. (2017). The effects of concordance-based electronic glosses on L2 vocabulary learning. *Language Learning & Technology, 21*(2), 32–51.
Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics, 28*(1), 46–65. https://doi.org/10.1093/applin/aml048.
Webb, S. (2008). The effects of context on incidental vocabulary learning. *Reading in a Foreign Language, 20*(2), 232–245.
Webb, S., Newton, J., & Chang, A. (2013). Incidental learning of collocation. *Language Learning, 63*(1), 91–120. https://doi.org/10.1111/j.1467-9922.2012.00729.x.
Wesche, M., & Paribakht, T. S. (1996). Assessing second language vocabulary knowledge: depth versus breadth. *Canadian Modern Language Review, 53,* 1–28. https://doi.org/10.3138/cmlr.53.1.13.
Xu, J. (2010). Using multimedia vocabulary annotations in L2 reading and listening activities. *Calico Journal, 27*(2), 311–327. https://doi.org/10.11139/cj.27.2.311-327.
Yanguas, I. (2009). Multimedia glosses and their effect on L2 text comprehension and vocabulary learning. *Language learning & Technology, 13*(2), 48–67.
Yoshii, M. (2006). L1 and L2 glosses: Their effects on incidental vocabulary learning. *Language Learning & Technology, 10*(3), 85–101.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.