THIS IS AN EARLIER VERSION OF THE MANUSCRIPT. FOR THE FINAL VERSION, PLEASE CHECK THE JOURNAL WEBSITE.

TESOL Quarterly: https://onlinelibrary.wiley.com/journal/15457249

The effect of equal versus expanding spacing practice on the deliberate learning of L2 collocations

Abstract

Research comparing equal and expanding spacing schedules in relation to L2 single words is limited. Moreover, none of the existing studies has examined the learning of formulaic language. The current study aims to address this gap by investigating the impact of equal and expanding intervals on the long-term retention of L2 collocations in deliberate learning contexts. Eighty-three university learners of English l (L1 Arabic) participated in the study. They were divided into three groups: equal spacing (7-7-7) condition, expanding spacing (1-5-15) condition, and a control condition (no exposure). The experimental groups completed four treatment sessions in which they were asked to deliberately learn 23 collocations. To assess collocation knowledge before and after the treatment, all groups took form recall and form recognition tests. Results show that both spacing conditions led to significantly higher gains than the control group in both tests. The difference between the two experimental groups was not significant for the recall measure but it was significant for the recognition measure (higher gains under 'equal spacing'). Moreover, estimated language proficiency modulated the difference between the control and expanding conditions, with a sharper difference as proficiency increased. Important implications for language teaching and classroom instruction based on these findings are discussed.

Keywords: vocabulary, input spacing, equal vs. expanding practice, collocations, deliberate learning

Introduction

Mastery of collocations, words that frequently co-occur with each other, is considered an essential component for achieving high levels of second language (L2) proficiency (Nation, 2013; Schmitt, 2022). Despite occupying a central role in second language learning, knowledge of collocations has been reported to develop more slowly than that of single words (e.g., Szudarski, 2017). Research shows that attaining collocational competence remains a challenge for many L2 learners, including those at advanced levels of proficiency (e.g., Nguyen & Webb, 2017; Siyanova-Chanturia, 2015). As a result, in the last couple of years, researchers and educators alike have examined different instructional methods to help learners increase their repertoire of collocations, including textual enhancement, direct instruction, and frequent exposure (e.g., Sonbul & Schmitt, 2013).

When a lexical item is frequently encountered in context, an important variable to take into consideration is the temporal distribution of these multiple exposures to the same target collocation. Studies suggest that distributed practice, in the form of either equal or expanding spacing schedules, has positive effects on vocabulary learning (Kang et al., 2014; Karpicke & Bauernschmidt, 2011; Nakata, 2015; Pyc & Rawson, 2007). However, this research has hitherto mainly been limited to single words only (though see Macis et al., 2021; 2023). Considering the importance of collocations and the learning difficulties they usually entail, it is relevant to explore how different encounters with the same collocation should be spaced to optimize learning. Another limitation of the available research studies is the fact that they have been conducted in either tightly controlled laboratory settings or computer-based classroom environments, which raises the question of the applicability of the findings to more traditional, teacher-directed classroom instruction. The present study thus aims to build on the limited existing vocabulary research and sets out to compare the effect of equal and expanding spacing on the long-term acquisition of L2 collocations in an authentic, teacher-led learning context.[1] The results of this research can potentially determine the effectiveness of different practice schedules and provide both learners and teachers with useful suggestions on how to improve instructed collocation learning.

Literature Review

Distributed Practice

The question of ‘distributed practice’ (also known as ‘input spacing’) is concerned with how the time intervals that elapse between repeated encounters can be manipulated to positively impact L2 performance (Cepeda et al., 2006; Logan & Balota, 2008). When discussing the effects of spacing, it is important to define two key phenomena. First, the ‘spacing effect’, which suggests that spaced gaps between repetitions result in better retention of knowledge than when all the repetitions are massed, with no or little time delay between them (Cepeda et al., 2006; Toppino & Gerbier, 2014). The second phenomenon is the ‘lag effect’, which refers to the differential impact of shorter versus longer time gaps (Rogers, 2017). The notion ‘distributed practice effect’ has been used as an umbrella term for both the 'spacing effect' and 'lag effect' (Cepeda et al., 2006). In the present study, we use distributed practice in referring to the 'spacing effect' but use 'lag effect' when contrasting the two spacing conditions, equal and expanding. The effect of time distribution has been extensively researched and evidenced in cognitive psychology (Cepeda et al., 2006). Moreover, several recent meta-analyses have also generally supported the robustness of distributed practice (e.g., Latimier et al., 2021)

Given the potential of distributed practice for optimizing L2 performance, the recent years have witnessed a surge of studies among SLA researchers. Spacing effects have been examined with respect to grammar learning (e.g., Kasprowicz et al., 2019; Serfaty & Serrano, 2022), vocabulary learning (e.g., Macis et al., 2021, 2023, Rogers & Cheung, 2021; Serrano & Huang 2021; Snoder, 2017; Yamagata et al., 2021) and general language skills (e.g., Serrano & Muñoz, 2007). When it comes to finding the optimal spacing condition, these empirical investigations have generally shown mixed results. In fact, the complex nature of distributed practice has been further emphasized by two recent meta-analyses that have specifically focused on second language learning (Kim & Webb, 2022; Uchihara et al., 2019). In the next section, we will review the evidence on equal versus expanding spacing in SLA research along with relevant theories proposed to explain the findings.

Theoretical Underpinning of the Lag Effect and Research Evidence

Three dominant theoretical accounts have been used to explain the mechanisms underlying the spacing effect. The first account – encoding variability theory – posits that encountering a target item at longer delays and in diverse contexts increases the richness of memory traces related to that item, which in turn results in more cues that can be used for successful retrieval (Benjamin & Tulis, 2010). The main premise of the second theory – deficient processing (Toppino & Gerbier, 2014) – is that encountering the target item too quickly after the initial encounter will result in diminished encoding because the first encounter may still be active in the short-term memory. When study sessions are more widely spaced, the initial presentation is no longer readily accessible in the short-term memory and more attention will need to be allocated to the subsequent repetition. Finally, study-phase retrieval theory (Toppino & Bloom, 2002; Toppino & Gerbier, 2014) holds that each subsequent repeated exposure to the target item after the initial one can only improve learning if a learner succeeds in reactivating (tracing) the original information about that item. If the repeated study sessions are wide apart, retrieval of the original information will be more effortful. In other words, longer spacing can result in 'desirable difficulties' (Bjork et al., 2013), which refer to the learning conditions that require more cognitive effort, thereby improving performance. Shorter spacing, on the other hand, does not yield as much effort and is therefore not optimal for retention of knowledge. In sum, according to the desirable difficulties framework, the amount of spacing that is neither too long nor too short results in desirably effortful retrieval that can lead to better L2 knowledge (e.g., Suzuki et al., 2020).

Several studies on distributed practice have explored the effects of 'relative spacing' or the 'lag effect', i.e., whether learning can be facilitated if the practice sessions are organized in an equal manner or if the intervals between them are gradually increased (Cull, 2000; Karpicke & Roediger, 2010; Storm et al., 2010). In equal spacing, the intervals between the learning episodes are kept uniform. For example, in the 3-3-3 schedules, learners study a target item and then encounter it afterwards three times in total, with these repetitions always being separated by three spacing units (e.g., days).[2] In the 1-3-5 schedules, on the other hand, the separation increases progressively, with the total treatment period or 'absolute spacing' remaining the same. The term 'absolute spacing' refers to the total amount of spacing that separates all the encounters of a target item (Karpicke & Bauernschmidt, 2011). In the example above, both schedules have the same absolute spacing of nine days (3-3-3 and 1-3-5) but differ in the relative distribution of practice sessions.

There have been quite a few studies examining equal and expanding spacing regarding linguistic materials (e.g., face-name associations) other than L2 vocabulary (e.g., Cull, 2000; Karpicke & Roediger, 2010; Storm et al., 2010). Surveying this literature is beyond the scope of this paper; however, because relative spacing seems to impact the learning of L2 vocabulary similarly to the learning of other materials (Nakata, 2015), we will briefly summarize the main conclusions that have been drawn from such research. In general, these studies show a mixed picture when it comes to the effect of equal and expanding spacing. Some of them found an advantage for expanding spacing (Karpicke & Roediger, 2007, Experiment 1; Logan & Balota, 2008), whereas others failed to do so (Cull, 2000; Karpicke & Roediger, 2010; Storm et al., 2010, Experiment 1).

The Lag Effect in L2 Vocabulary Research

L2 vocabulary research examining whether study sessions should be spaced in an equal manner, or whether gradually increasing practice sessions may be more beneficial for learning is also mixed. Moreover, there is far less research on second language vocabulary learning than on other linguistic aspects. Kang at al. (2014), for example, used word pairs as the learning material for their experiment with Japanese learners and implemented days as an index of spacing. In a four-week long experiment, 37 learners of English were instructed to study 60 Japanese-English word pairs following either an equal spacing schedule (9-9-9 day) or an expanding spacing schedule (2-6-19 day). There were no statistically significant differences between the two spacing schedules on the self-paced L1-L2 translation delayed posttest administered 53 days after the end of the treatment (equal=46%, expanding=49%).

In three other studies on the lag effect on vocabulary learning (Karpicke & Bauernschmidt, 2011; Nakata, 2015; Pyc & Rawson, 2007), trial, rather than day, was used as a unit of spacing whereby target items were separated by other intervening words in the same session. In Pyc and Rawson (2007), 161 participants studied 24 Swahili-English word pairs under either equal (5-5-5) or expanding (1-5-9) spacing schedules. The treatment was conducted in one session and the learning was measured in a self-paced cued-recall test that was administered 40 minutes after the treatment. Findings indicate 67.6% of correct recall for equal and 63.2% of correct responses for expanding spacing and no statistically significant differences between the two conditions. Similarly, in Karpicke and Bauernschmidt (2011), 96 participants were asked to study 24 Swahili-English word pairs. The authors compared six different spacing conditions: 5-5-5 (short equal), 1-5-9 (short expanding), 10-10-10 (medium equal), 5-10-15 (medium expanding), 30-30-30 (long equal) and 15-30-45 (long expanding). Retention was measured through a one-week L2-L1 translation delayed posttest. Results show that long spacing led to 75% of correct responses. This was followed by the medium spacing conditions that resulted in gains of 64% and the short spacing conditions that produced 49% of correct responses in the test. Finally, Nakata (2015) investigated whether the amount of spacing (massed, short, medium, long) and the type of spacing (equal and expanding) had any differential effects on the learning of 20 Japanese-English word pairs. The author compared the same spacing conditions as in Karpicke and Bauernschmidt (2011). The treatment was conducted in one session and the learning gains were measured immediately and one week after the session by means of two tests, one receptive (L2-L1 translation) and one productive (L1-L2 translation). As opposed to the previous L2 vocabulary research, Nakata found a significant advantage (albeit small) of expanding spacing over equal on the receptive delayed post-test when gains were collapsed across the short, medium, and long spacing conditions. The gains were 64.3% for expanding and 61.5% for equal spacing. The differential results may be due to the type of task used during the treatment (for example, Nakata used productive retrieval whereas other studies used receptive retrieval) as well as the differences in the lengths of the lags used in these studies.

Although the results of the above studies are informative, they suffer from at least two limitations. Firstly, in most of them spacing was manipulated within a single session, in laboratories or in computer-based learning contexts. This raises the question of generalizability to common classroom environments where material is usually reviewed across several classes and the classes are led by a teacher. Secondly, the available research has only focused on single words. It is now widely attested that the English language is made up of large portions of fixed and semi-fixed lexical units that play a fundamental role in language learning and use (Siyanova-Chanturia & Pellicer-Sánchez, 2020). Phrasal verbs, idioms, and more specifically, collocations, are not only widespread in the language but also represent one of the greatest hurdles to native-like fluency (Schmitt, 2022). Thus, one key pedagogical concern is what instructional methods can assist learners in overcoming these hurdles as quickly as possible. The next section focuses on one type of these lexical units, namely collocations.

Learning of L2 Collocations

At the outset, it is important to define collocations. The present study adopts the frequency-based approach according to which collocations are operationalized as combinations of two words which co-occur with each other with frequency higher than chance, irrespective of their semantic transparency (see Barfield & Gyllstad, 2009; for more on approaches to defining collocations). Thus, in the present study, semantic transparency was only considered as a covariate in the analysis to partial out its effect on gains (see *Materials* and *Analysis* below). Research has shown that L2 learners' knowledge of collocations is often limited, and that the development of collocational knowledge tends to progress at a slow rate, even for highly proficient L2 learners (e.g., Laufer & Waldman, 2011). A number of studies have indicated that the challenges learners experience can be due to the compositional nature of some collocations (e.g., Macis & Schmitt, 2017; Peters, 2014), lack of awareness that the two lexical items can form a larger unit of meaning (e.g., Laufer, 2011), negative L1 influence (e.g., Wolter & Gyllstad, 2011), and insufficient L2 exposure (e.g., Boers, 2020).

More importantly, previous research also suggests that the retention rates of collocations may be affected by whether learners study them intentionally or whether collocations are learned incidentally, as a by-product of another task such as general comprehension (Hulstijn, 2003). Several empirical investigations have demonstrated that deliberate learning (the focus of this study) is generally an effective approach in promoting collocational knowledge (Chan & Liou, 2005; Sonbul & Schmitt, 2013; Tsai, 2020), and that it results in greater and faster learning gains when compared to learning collocations incidentally (Laufer & Girsai, 2008; Szudarski, 2012).[3]

In an attempt to examine the relationship between collocational learning condition (incidental vs. deliberate) and distributed practice, Macis et al. (2021)

compared the effects of spacing and massing on the learning of 25 adjective-noun collocations in incidental (Experiment 1) and deliberate (Experiment 2) learning situations. The participants in the incidental group read short stories that were followed by comprehension questions. In contrast, those in the deliberate group learned the target collocations through concordance lines and then completed matching and multiple-choice exercises. Over a period of five weeks, each target collocation was encountered five times. In the massed groups, target collocations were encountered five times on the same day whereas in the spaced groups, they were encountered once a week. Results showed that, when collocations were learned incidentally, massing was more effective. More importantly, when the participants engaged in the deliberate learning of collocations, spacing led to much better results. Findings also indicate that the participants' proficiency, as measured by the Updated Vocabulary Test (Webb et al., 2017), was shown to be a significant predictor of their learning gains. Although Macis et al. (2021) is, to the best of our knowledge, the only study on lexical distributed practice that employed an objective measure of proficiency, it did not aim to evaluate the potential modulating effect of proficiency on lexical gains under the various conditions. This is an important factor to consider as proficiency emerged as a significant modulator (of frequency and transparency effects) in studies that examined L2 collocation processing (e.g., Sonbul, 2015; Sonbul & El-Dakhs, 2020; Wolter & Gyllstad, 2013; Yamashita & Jiang, 2010). Another limitation of Macis et al. (2021, particularly Experiment 2) is that the authors only looked at equal spacing. Given the potential positive effect of expanding spacing on the learning of single words (Nakata, 2015, see above), it would be useful to explore whether the positive effects of this practice schedule also extend to collocations. To the best of our knowledge, no study to date examined the optimal spacing condition for learning collocations deliberately.

The Present Study

The present study is a follow-up on the Macis et al. (2021) study. Specifically, we examined the effectiveness of two spacing conditions during deliberate learning sessions: equal (7-7-7 days) and expanding (1-5-15 days). Thus, the total treatment period was the same for both conditions (i.e., 21 days), the only difference being in the gaps between the sessions (see Figure 1). We also included a control group as a baseline for comparison.

[insert Figure 1 around here]

The rationale for the chosen spacing conditions was partially pedagogical. The 7-7-7 scheme would mirror a common instructed learning lesson frequency of one lesson per week. In settings where there is little input for learners outside of taught school sessions, the weekly lesson may be the only opportunity for them to get exposure to the target language. The 7-7-7 equal spacing corresponds to a condition in between what Karpicke and Bauernschmidt (2011) called 'short' and 'medium' conditions. Since we wanted to keep the total treatment period the same for our experimental groups, the expanding spacing scheme in the present study (similar to Kang et al.'s, 2014, expanding condition and corresponding to 'medium' in Karpicke and Bauernschmidt's, 2011, taxonymy) was a consequence of this. It should also be noted that the expanding spacing condition, as operationalized in the present study, reflects a situation where the first repetition comes very close to the initial exposure, and a second repetition earlier than the equal spacing group had their first repetition.

The study adopted a quasi-experimental design (pretest – treatment – delayed posttest) and tested two levels of knowledge before and after treatment: form recall and form recognition. Deliberate learning is operationalized in the present study (following Macis et al., 2021) as memorizing target collocations, embedded in short concordance lines, and completing follow-up exercises. The following research questions were addressed:

RQ1: To what extent do deliberate learning conditions (with expanding or equal spacing) lead to long-term form-recognition and form-recall collocation knowledge gains?

RQ2: Is one of the two deliberate learning conditions (with expanding or equal spacing) more effective in developing long-term form-recognition and form-recall collocation knowledge gains?

To answer RQ1, we compared gains (form recall and form recognition) under the two experimental learning conditions to the scores of the control group; and to answer RQ2, we examined the differential gains under the two experimental conditions. We additionally employed a rough measure of proficiency (vocabulary size test) to examine any potential modulating effect of proficiency on variation in gains among conditions, an aspect generally ignored in earlier distributed practice vocabulary research.

Methods

Participants

Initially, six intact classes took part in the study. Students in these classes were first-year L1 Arabic female students at a university in Saudi Arabia. They were studying a preparatory B1 (pre-intermediate) English language course. Classes were randomly assigned to one of the three conditions: expanding spacing, equal spacing, and control. The purpose of the study was explained to the students, and those who were willing to participate signed the informed consent. Only data from students who attended all testing and treatment sessions was included in the analysis.

Thus, we ended up with 83 participants (equal spacing, $n = 27$, expanding spacing, $n = 25$, and control, $n = 31$). They ranged in age between 17 and 20 ($M = 18.30$, $SD = 0.56$) and started learning English at an average age of 11 ($M = 11.04$, $SD = 2.76$). The participants took the V_YesNo online vocabulary test (Meara & Miralpeix, 2017; maximum score = 10,000) as a proxy of general proficiency. Their average score was 2,749.16 ($SD = 1158.27$). This roughly indicates knowledge of the most frequent 2,000 word families in English. We conducted an ANOVA to compare the vocabulary scores of the students under the three conditions and found no significant differences among them (*equal spacing:* $M = 2{,}704.7$, $SD = 1086.5$, *expanding spacing:* $M = 2{,}723$, $SD = 907.95$, *control:* $M = 2{,}808.97$, $SD = 1362.09$, $F(2) = 1.57$, $p = .21$).[4] We included the vocabulary score as a main and interacting variable in the mixed-effects models (see *Analysis*) to examine the effect of approximate proficiency on posttest scores and to inspect any modulating effect of proficiency on the differences between conditions.

Materials

The collocation items were drawn from the study by Macis et al. (2021). Macis et al. developed a set of 25 adjective-noun collocations which were (1) highly frequent (50 occurrences in the COCA and 10 in the BNC), (2) with a high mutual information (MI) score of 3+, (3) incongruent with Arabic, and (4) could be elicited from natives in a cued gap-fill test (see '*Measures*' below). The items were rated for transparency on a scale from 1 (opaque) to 7 (transparent). Full details of congruency and transparency operationalization are included in Macis et al. (2021).

In the present study, we used 23 out of the total 25 items included in Macis et al. (2021). As we developed a form recognition, multiple-choice (MC), test in addition to the form recall, gap-fill, test used by Macis et al., it was not possible to develop MC items for two of the original collocations (*final leg* and *poor taste*) following the strict criteria that were employed (see '*Measures*' below and Appendix S3). The full list of 23 items is provided in Appendix S1 (Supplementary Materials).

The materials for the four treatment sessions were identical for both experimental groups and were also borrowed from Macis et al. (2021, Experiment 2). In each session, the participants were presented with the 23 target items within concordance lines. There was also a task to assess learning at the end of each session. The tasks used in Macis et al.'s study were all receptive (MC and matching), but the tasks we developed for the present study ranged between receptive and productive. This was because our two measures were both receptive (recognition) and productive (recall), and we intended to ensure balanced practice for both types of knowledge. The full materials are included in Appendix S2 (Supplementary Materials). It should be noted that Macis et al.' study included five treatment sessions, but in the present study we only included four due to the academic calendar restrictions.

Measures

Form recall

To assess form-recall knowledge of the target collocations, we used the same test developed by Macis et al. (2021). The format was a cued gap-filling where the noun node of each target collocation was presented in a defining context and the adjective was removed. The participants were instructed to provide the missing adjective for each collocation. To restrict responses to the target adjectives, two types of clues were included: the first letter(s) and spaces indicating the number of missing letters. The test was validated with ten native speakers, and 80% provided the target response. Here is an example of form recall item for the target 'broken promises':

The people are tired of all the government's b _ _ _ _ _ promises.

The vocabulary in the context belonged to the most frequent 2,000 word families in English (Nation, 2012) to match the learners' level (see *Participants* above). Any words that belonged to lower frequency levels were translated in the margin to ensure maximum comprehension.

The responses in the form recall test were awarded a binary score of 0 (incorrect) /1 (correct) by a research assistant who held an MA degree in English/Arabic translation. A response that was correct in spelling and morphology was scored 1. A recognizable response was also awarded a score of 1 regardless of minor spelling mistakes (e.g., ‘notural’ for ‘natural’) or morphological errors (e.g., providing a noun in place of an adjective, e.g., ‘pain’ for ‘painful’). A clearly incorrect response or a missing answer received 0. As the research assistant started scoring the responses, she discussed several of the responses with the first author until an agreement was reached. Once the scoring was complete, and to check the reliability of the scoring procedure, 33% of the responses were checked by the first author. As the inter-rater reliability was high (ICC = 0.96), only the scores awarded by the research assistant were included in the analysis.

Form recognition

To assess knowledge at the form recognition level (in addition to form recall), we developed an MC test. The same sentence contexts included in the form recall test were used in this test, albeit without clues. The order of items was altered to reduce the test-retest effect. The target adjective was missing, and five choices were presented for each (the key, three distractors that did not collocate with the noun, and an 'I don't know' option to reduce guessing). Here is an example for the collocation 'broken promises':

The people are tired of all the government's __________ ***promises****.*

A. *bad*

B. *broken*

C. *damaged*

D. *undone*

E. *‘I don’t know’*

The criteria for selecting distractors and their specified properties are included in Appendix S3 (Supplementary Materials). To validate the form recognition test, we administered it to 15 native speakers of English who showed a perfect score. The final form recognition test is presented in Appendix S4 (Supplementary Materials). Scoring of the test was straightforward. A correct answer was awarded a score of 1 whereas any incorrect answer, missing response, or an 'I don't know' option received 0.

Procedures

The study was conducted in six sessions: pre-testing session, four treatment sessions for the experimental groups, and a delayed post-testing session. First, all participants under all conditions (equal spacing, expanding spacing, and control) completed the form recall and form recognition pretests. The two measures were carefully sequenced to avoid any effect from one test on the other. The participants were administered the form recall test first which was collected upon completion. Then, we administered the form recognition test and ended the session with the V_YesNo online vocabulary test as a proxy measure of proficiency.

Two weeks later, the treatment started under the two experimental conditions, and it spanned over a period of 21 days. The first treatment session (Session 1) was conducted on the same day for both conditions. Then, we followed a strict spacing schedule for each group according to their designated condition. Participants under equal spacing completed one treatment session per week while those under the expanding spacing condition followed a 1-5-15-day spacing schedule. Although the spacing between the sessions for the two experimental conditions was different, the absolute spacing period was identical for both. The participants in the control condition followed their regular classes and did not receive any treatment.

Participants under both experimental conditions started the treatment (Session 1) on the same day and finished on the same day (Session 4). Moreover, the procedures within each session were the same for both conditions. At the beginning of each treatment session, each participant was given a handout which included the full list of 23 target collocations in context. The instructions explicitly stated that the collocations needed to be committed to memory for an upcoming task, and that it would not be possible to refer back to the handout to complete that task. Ten minutes later, the sheets were collected, and the task was administered. Participants were given 5 minutes to complete each task, and no feedback was provided after the task to avoid any extra exposure to the target items. Each treatment session took 15 minutes to complete in total. The scores achieved by participants in the experimental groups are presented in Appendix S5, showing no difference in scores in Sessions 1 and 2 but a clear advantage for the equal spacing group in Sessions 3 and 4 and in the average score ($p < .001$).

Two weeks after the last treatment sessions, all participants (including those in the control group) took the delayed post-tests administered in the same order as in the pre-testing session. At the end, the participants were asked to complete a short language background questionnaire.

Analysis

Initially, we inspected the internal consistency of the two measures in the pre- and post-testing sessions and found it to be acceptable: Form recall (pre-test Cronbach alpha = .68, post-test Cronbach alpha = .95) and form recognition (pre-test Cronbach alpha = .81, post-test Cronbach alpha = .91).

The analysis was done in R version 4.1.1 (R Development Core Team, 2021). We employed mixed effects logistic regression models for binary data (correct = 1, incorrect = 0) using the *glmer* function with a binomial family under the *lme4* package, and two models were fit. The model structure was the same in both cases, with the only difference being the dependent variable: form recall scores in Model 1 and form recognition scores in Model 2. Categorical predictors were dummy coded. We followed the stepwise forward method of modelling whereby predictors were added one by one to the null model, and their contribution was inspected based on AIC values and significance level. Only predictors that significantly improved the model fit (reducing the AIC value at $p < .05$) were kept until we reached the final best-fit model for each measure. The detailed model structure with a summary of continuous variables are presented in Appendix S6.

We used exponentiated regression coefficients, Exp (β)), transformed from log odds as predictors of the odds of a correct answer for each variable. We further calculated *Cohen's d* as a standardized measure of effect size, and interpreted the values based on Plonsky and Oswald’s (2014) guidelines for between-subject comparisons in L2 research: small (0.40), medium (0.70), and large (1.00). Marginal and conditional R^2 values were also computed for each model using the *rsq.glmm* function in the *rsq* package.

Results

Raw gains

Table 1 presents the number and percentage of correct answers in both testing sessions under all conditions. The percentage of correct answers in the pretest were similar across the three conditions in both measures. Moreover, all conditions showed a certain level of improvement in the posttest, which is an expected test-retest effect.

Looking at the form recall measure, we can see that the control group showed an improvement of 7.4 percentile points in the posttest over the pretest. We took this figure (7.4) as the baseline level against which we compared form-recall gains under the two experimental conditions. Correct responses under the equal spacing condition were 60.9 percentile points over this baseline or around 14 collocations (out of the total 23 targets). For the expanding spacing condition, the increase over the control group was approximately 56.2 percentile points (i.e., 13 collocations). Thus, in terms of raw gains, the equal spacing condition led to the learning of one additional collocation over the expanding spacing condition.

As for the form-recognition scores, one can observe higher scores overall in comparison to form recall. This is expected given that receptive knowledge is easier to develop than productive knowledge. The baseline posttest – pretest score under the control condition was around 8 percentile points. Comparing this baseline to the gains of the two experimental conditions, we observe 43.2 percentile points for the equal spacing condition and 40 percentile points for the expanding spacing condition (10 collocations and 9 collocations learned over the control group, respectively). Therefore, similarly to the form-recall scores, the equal spacing condition showed a one-collocation increase in raw gains over the expanding spacing condition.

[insert Table 1 around here]

Form recall gains (Model 1)

Table 2 presents the best-fit model for variables predicting form-recall scores. First, it shows that as the first word length was increased by one unit (i.e., when the adjective is one letter longer), the odds for a correct response were decreased by 50% (1/0.64). Second, a one unit increase in log vocabulary score (an increase of 10 points out of 10,000) was associated with an over two times increase in the odds of a correct response. Pretest scores had a clearly large effect on posttest scores whereby items known in the pretest were almost 26 times more likely to be known in the posttest.

Looking at the main effect of Condition, we can clearly see that both experimental conditions led to significantly higher scores in the posttest than the control condition with a large effect. In comparison to the control group, participants under the equal spacing and the expanding spacing conditions were almost 74 times and over 50 times more likely to score correctly in the posttest, respectively. We redefined the reference level to 'equal' in order to inspect the third comparison between the 'equal' and 'expanding' conditions, and the difference was not significant (*estimate* = -0.38, *z value* = -1.00, *p* = .32, *Exp* = .68, *Cohen's d* = -0.21). Finally, it is notable that the interaction between vocabulary score and Condition did not significantly contribute to the form-recall model fit.

[insert Table 2 around here]

Form recognition gains (Model 2)

The best-fit model for scores in the form recognition test is presented in Table 3. First, the length of the second word (the noun node) significantly contributed to the model fit with a 1-letter increase being associated with 30% (1/0.77) decrease in the odds of a correct response. Then, a one unit increase in the log vocabulary score was associated with 92% increase in the odds of a correct MC response. The pretest scores also significantly improved the model fit with items receiving correct responses in the pretest being three times more likely to receive a correct response in the posttest. Finally, both treatment conditions led to significantly higher scores than the control group with a large effect (the odds being over 40 times and over 13 times for equal and expanding spacing, respectively). The difference between equal and expanding spacing was also significant with a small effect (*estimate* = -1.13, *z value* = -2.46, *p* = .01, *Exp* = 0.32, *Cohen's d* = -0.62). The odds of a correct posttest response under the equal spacing condition were over three times (1/0.32) larger than the odds under the expanding spacing condition.

[insert Table 3 around here]

Finally, the interaction between Condition and vocabulary score was significant, with vocabulary scores significantly modulating the difference between the control and expanding spacing conditions. When the reference level was redefined to 'equal', it was found that the vocabulary score did not have any significant modulating effect on the difference between the 'equal' and 'expanding' conditions (*estimate* = 0.01, *z value* = 0.02, *p* = .98, *Exp* = 1.01, *Cohen's d* = 0.01). The effect is graphically presented in Figure 2. We can clearly see that the control condition always led to lower scores than the experimental conditions, but the difference was clearer and less variant (see confidence intervals) as proficiency increased. Based on the significance values above, it should be noted that this modulating effect of proficiency only reliably applied to the difference between ‘control’ and ‘expanding’ conditions. Figure 2 also shows that the difference in probabilities between the equal and expanding conditions was clearer towards the lower side of proficiency. This difference shrank at the highest levels of proficiency, albeit with large variation, hence the insignificant effect.

[insert Figure 2 around here]

Discussion

The present study investigated deliberate learning of English collocations by L1 Arabic university-level students. A between-subject quasi-experimental design with two experimental groups and one control group was used and two measures of knowledge were employed (form recall and form recognition). The aim was to evaluate the effectiveness of two spacing conditions, equal and expanding, over the control (baseline) condition (RQ1) and to determine whether one of the spacing conditions was more effective than the other (RQ2).

To answer RQ1, our results indicate that both deliberate learning conditions rendered clearly significant gains with large effect sizes compared to the control group. This finding is in line with previous research on the effectiveness of deliberate approach when learning L2 collocations (e.g., Laufer & Girsai, 2008; Sonbul & Schmitt, 2013). It should be mentioned that the control group in the present study experienced no alternative collocation-oriented treatment but attended their normal lessons. A future alternative could entail a condition tapping into incidental learning. This should help weigh the relative effectiveness of different spacing schemes under diverse exposure conditions (Macis et al., 2021).

Moreover, it is worth considering the magnitude of learning in the experimental groups in light of the short treatment sessions. For form recognition, the experimental groups learnt 9-10 more collocations than the control group; and for form recall, they outperformed the controls by 13-14 collocations. It should be noted, then, that each learning session was short, only 10 minutes for learning followed by 5 minutes for the task part. Thus, a total of 60 minutes (4 sessions x 15 minutes) on task yielded a respectable uptake in both knowledge modalities. The gains in form recall for the equal spacing group in the present study are similar to those reported in Macis et al. (2021, Experiment 2), albeit higher (60.9 and 37.9 percentile-point advantage over the control group, respectively). The difference can be attributed to the timing of the post-test which was scheduled three weeks after the last treatment session in Macis et al. but two weeks post treatment in the present study. According to Schmitt (2010, pp. 156–157), the minimum gap between treatment and a measure of long-term gains is two days to capture "the integration of a new lexical item into the mental lexicon", but, Schmitt adds, "three weeks should be indicative of learning which is stable and durable." Had our test been placed a week later, our learners might have experienced similar levels of attrition to Macis et al. (2021). More research on the lag effect should include a three-week gap prior to the post-test for a better understanding of long-term gains.

In relation to English proficiency, operationalized as scores on the V_YesNo online vocabulary test (Meara & Miralpeix, 2017), our results show that this variable modulated the effect on form recognition for lower-level learners, whereby the expanding spacing did not lead to higher gains over the control group (see Figure 2). Moreover, generally, learners with larger vocabularies tended to learn more vocabulary, a phenomenon referred to as the Matthew Effect: "the rich get richer" (see Webb & Nation, 2017). The assumed reasons for this have been suggested to stem from a better ability to focus more cognitive attention on the lower number of unknown words appearing in their input coupled with a reduced learning burden compared to lower-proficiency learners.

In RQ2 we asked whether one of the two experimental deliberate learning conditions was more effective in developing long-term form-recognition and form-recall collocation knowledge gains. Our results indicate that the two conditions were equally effective in developing form recall knowledge (gap-fill performance) but equal spacing was more effective in developing form recognition knowledge (MC performance), with just one more collocation learned on average. The form recall findings seem to support previous research which targeted individual words and showed no difference between the two spacing conditions (Kang et al., 2014; Karpicke & Bauernschmidt, 2011; Pyc & Rawson, 2007). However, our form recognition findings stand in sharp contrast with those of Nakata (2015) where exactly the opposite effect was reported: an advantage for expanding spacing. This might either be attributed to differences in the operationalization of spacing (intervening days vs. items), the measure employed (MC vs. translation), the delay (two weeks vs. one week), or the nature of the target lexical item (individual words vs. collocations). Moreover, the results of both form recall and form recognition posttests cannot be fully accounted by any of the theoretical accounts discussed in the literature review (*encoding variability*, *deficient processing*, or *study-phase retrieval*). All of these frameworks essentially state that longer delays between encounters will lead to more learning, which was in contrast to the findings of the present investigation. More research is needed to tease these factors apart. Future research in this area can compare individual words to collocations (cf. Laufer & Girsai, 2008) by employing different measures of knowledge and operationalizing spacing in term of days (rather than items) for findings that have clear pedagogical implications.

It is also intriguing to see that we found an advantage for equal spacing only under form recognition but not under from recall. This can potentially be attributed to the level of knowledge examined. Form recognition is considered less demanding than form recall (Laufer & Goldstein, 2004), as the latter involves an activation, retrieval, of the form of a collocation when prompted by the collocation's meaning. Thus, arguably, form recall was far too difficult to develop for all learners (regardless of exposure condition) and thus did not show any advantage for one spacing scheme over the other. This difference between recall and recognition results should, however, be interpreted with caution as the trend is the same for both measures, with one out of 23 collocations better retained under equal spacing over expanding spacing.

The relative advantage for equal over expanding spacing might seem counter intuitive. It is tempting to think that the way repetitions were sequenced in the present study (with a first repetition already the day after the initial learning session took place for the expanding condition, see Figure 1) would be facilitative compared to the equal spacing condition which had its first repetition a full week after the initial session. In contrast, by the time the equal spacing provided its first repetition, the expanding spacing group had already been subjected to their second repetition. How can the latter be explained?

The equal spacing advantage for form recognition can potentially be explained to some extent by a 'desirable difficulties framework' (Bjork et al., 2013). According to the framework, each subsequent repeated exposure to a target item after the initial one can only improve learning if a learner succeeds in reactivating (tracing) the original information about that item. What seems to follow from this is that a too long gap between the first and the second exposure is negative for learning. At the same time, if the repeated study sessions are somewhat wide apart, then retrieval of the original information will be more effortful, leading to better learning. The scores of our participants in the post-treatment task (administered at the end of each session, see Appendix S5) seem to support this argument. Learners under both conditions performed similarly in the first two tasks (Session 1 and Session 2) but the equal spacing group exhibited a clear advantage in Session 3 and Session 4. Thus, one might argue that with longer gaps under the expanding spacing condition, learners' memory of the target items diminished from session to session, especially that no feedback was given at the end of each session which should have boosted the effect of treatment through participants' verifying and rectifying their responses. This arguably led to more cognitive effort in retrieving earlier exposures to the target item and in committing them to memory under the expanding condition in the later sessions as learners (particularly the less proficient ones) needed to re-study the target collocations. Thus, spacing that is neither too short nor too long (the equal spacing treatment schedule with one session per week) results in desirably effortful retrieval, which facilitates better knowledge (Suzuki et al., 2020). It should be noted here that results of the form recognition test in the present study should be interpreted cautiously for two reasons. First, the reported advantage for the equal spacing condition accounts for only one collocation over the expanding spacing condition. This is too small to make strong arguments regarding the pedagogical value of one spacing schedule over the other. Second, a certain level of test-order effect on recognition scores (with a preceding cued recall test) is inevitable. We have attempted to control for this effect through the inclusion of a no-exposure group who took all tests in the same order and thus represented baseline performance against which reliable gains were established. However, one may argue that test-order effects might vary between the control and experimental groups due to exposure. This is an empirical question that can be addressed in future research.

Our study has several implications for the teaching and learning of collocations in classrooms. First, our results highlight the value of deliberate learning of collocations (Laufer & Girsai, 2008; Sonbul & Schmitt, 2013), which resulted in durable gains under both spacing conditions. Thus, although the literature often argues that collocations are best learned incidentally through rich exposure (e.g., Nation, 2013), this should not make deliberate learning of highly frequent collocations irrelevant. Then, given that the two spacing conditions gave rise to quite similar outcomes, the equal spacing scheme (akin to our study's 7-7-7 spacing), which provides a "not too late or too soon" approach, is encouraged as it is easier for teachers and learners to follow. This is especially the case for lower-level L2 learners who might struggle to develop form recognition knowledge when the repetitions are scheduled in an expanding fashion; unlike 'stronger' L2 learners who can successfully cope with relatively long gaps. This variation in learners' response to spacing schemes additionally implies that textbook developers need to carefully consider the spacing option that is most suitable for different learner groups. In early stages, target collocations should be spaced at equal intervals (e.g., once per chapter) to avoid the potential attrition associated with expanding gaps. As the textbook series progresses in levels, a mixture of spacing options (equal and expanding) can be implemented as higher-level learners can be more resilient against longer gaps between repetitions. In mixed-level groups, it should be the role of the teacher to try to customize repetition intervals for different learners to account for individual differences in proficiency among them. This can potentially be done through encouraging learners to revise collocations outside the classroom following the schedule that best suits their proficiency level.

We hope this study will open the door for more research on optimal spacing schedules for L2 collocation knowledge learning. We suggest future studies comparing equal and expanding spacing to both replicate current findings and investigate alternative spacing schedules. In either type, keeping the absolute spacing period the same as well as the number of repetitions is paramount. Furthermore, controlling for learners' proficiency levels is essential. It would also be interesting to examine whether and how the results would change if feedback was provided after the treatment task and if the spacing schedule was different. For example, shorter gaps between sessions might arguably strike a better balance between retrieval effort and retrieval success. Another factor to consider is plausible exposure to target items outside the treatment sessions, especially for the experimental groups with an explicit focus on target items. This line of research should potentially help textbook writers and L2 teachers ensure the best conditions for vocabulary learning, including collocations, in the classroom environment.

Notes

1- Long term acquisition refers to learning that is measured by means of a delayed posttest at least a week after the treatment (Schmitt, 2010).

2- It is important to note that these numbers (e.g., 3-3-3) refer to the retrieval opportunities after the first exposure.

3- Studies have shown that the effectiveness of various explicit tasks will inevitably also depend on the degree of involvement with unknown vocabulary (Keating, 2008).

4- We used the *aov* function in the *stats* package to conduct the ANOVA. The vocabulary scores were log transformed and centered prior to the ANOVA, and the Shapiro-Wilk test showed normal distribution (p = 0.32).

References

Barfield, A., & Gyllstad, H. (2009). Introduction: Researching L2 collocation knowledge and development. In A. Barfield & H. Gyllstad (Eds), *Researching collocations in another language: Multiple interpretations* (pp. 1-18). Palgrave Macmillan.

Benjamin, A.S., & Tullis, J. (2010). What makes distributed practice effective? *Cognitive Psychology, 61*(3), 228-247.

Boers, F. (2020). Learning single words vs. multiword items. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 158-173). Routledge.

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology*, *64*(1), 417-444.

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

Chan, T.P., & Liou, H.C. (2005). Effects of web-based concordancing instruction on EFL students' learning of verb–noun collocations. *Computer Assisted Language Learning*, *18*(3), 231-251.

Cull, W. L. (2000). Untangling the benefits of multiple study opportunities and repeated testing for cued recall. *Applied Cognitive Psychology*, 14, 215–235.

Hulstijn, J. H. (2003). Incidental and intentional learning. In C. Doughty & M. H.

Long (Eds.), *The handbook of second language acquisition* (pp. 349-381). Oxford, England: Blackwell.

Kang, S. H. K., 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.

Karpicke, J. D., & Bauernschmidt , A. (2011). Spaced retrieval: Absolute spacing enhances learning regardless of relative spacing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(5), 1250-1257.

Karpicke, J. D., & Roediger, H. L. (2007). Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances long-term retention. Journal of Experimental Psychology: *Learning, Memory, and Cognition*, 33, 704-719.

Karpicke, J. D., & Roediger, H. L. (2010). Is expanding retrieval a superior method for learning text materials? *Memory and Cognition*, 38, 116-124.

Kasprowicz, R. E., Marsden, E., & Sephton, N. (2019). Investigating distribution of practice effects for the learning of foreign language verb morphology in the young learner classroom. *The Modern Language Journal*, *103*(3), 580-606.

Keating, G. D. (2008). Task effectiveness and word learning in a second language: The involvement load hypothesis on trial. *Language Teaching Research*, *12*(3), 365-386.

Kim, S. K., & Webb, S. (2022). The effects of spaced practice on second language learning: A meta-analysis. *Language Learning, 72*(1), 269-319.

Laufer B. (2011). The contribution of dictionary use to the production and retention of collocations in a second language. *International Journal of Lexicography*, 24, 29-49.

Laufer, B., & Girsai, N. (2008). Form-focused instruction in second language vocabulary learning: A case for contrastive analysis and translation. *Applied Linguistics, 29*(4), 694-716.

Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning*, *54*(3), 399-436.

Laufer, B., & Waldman, T. (2011). Verb-noun collocations in second-language writing: A corpus analysis of learners' English. *Language Learning, 61*(2), 647-672.

Latimier, A., Peyre, H., & Ramus, F. (2021). A meta-analytic review of the benefit of spacing out retrieval practice episodes on retention. *Educational Psychology Review, 33*(3), 959-987.

Logan, J. M., & Balota, D. A. (2008). Expanded vs. equal interval spaced retrieval practice: Exploring different schedules of spacing and retention interval in younger and older adults. *Aging, Neuropsychology, and Cognition,15*(3), 257-280.

Macis, M., Alharbi, R., & Sonbul, S. (2023). Investigating optimal spacing schedules for incidental acquisition of L2 collocations. *International Review of Applied Linguistics in Language Teaching (IRAL).*

Macis, M. & Schmitt, N. (2017). Not just small potatoes: Knowledge of the idiomatic meanings of collocations. *Language Teaching Research*, *21*(3), 321-340.

Macis, M., Sonbul, S., & Alharbi, R. (2021). The effect of spacing on incidental and deliberate learning of L2 collocations. *System*, *103*, 102649.

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

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.

Nation, I. S. P. (2012). The BNC/COCA word family lists. Document bundled with Range program with BNC/COCA lists, 25. Available online at: https://www.victoria.ac.nz/lals/about/staff/paul-nation#publications.

Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge, UK: Cambridge University Press.

Nguyen, T. M. H., & Webb, S. (2017). Examining second language receptive knowledge of collocation and factors that affect learning. *Language Teaching Research, 21*(3), 298-320.

Peters, E. (2014). The effects of repetition and time of post-test administration on EFL learners' form recall of single words and collocations. *Language Teaching Research, 18*(1), 75-94.

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning, 64*(4), 878-912.

Pyc, M. A., & Rawson, K. A. (2007). Examining the efficiency of schedules of distributed retrieval practice. *Memory & Cognition*, 35, 1917-1927.

R Core Team (2021). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org.

Rogers, J. (2017). The spacing effect and its relevance to second language acquisition. *Applied Linguistics, 38*(6), 906-911.

Rogers, J., & Cheung, A. (2021). Does it matter when you review? Input spacing, ecological validity, and the learning of L2 vocabulary. Studies in Second *Language Acquisition, 43*(5), 1138-1156.

Schmitt, N. (Ed.). (2004). *Formulaic sequences*. Amsterdam, Philadelphia: John Benjamins.

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Palgrave Macmillan.

Schmitt, N (2022). Norbert Schmitt's essential bookshelf: Formulaic language. *Language Teaching*, 1-12.

Serfaty, J., & Serrano, R. (2022). Lag effects in grammar learning: A desirable difficulties perspective. *Applied Psycholinguistics, 43*(3), 513-550.

Serrano, R., & Huang, H.-Y. (2021). Time distribution and intentional vocabulary learning through repeated reading: A partial replication and extension. *Language Awareness*, *32*(3), 1-19.

Serrano, R., & Muñoz, C. (2007). Same hours, different time distribution: Any difference in EFL? *System, 35*(3), 305-321.

Siyanova-Chanturia, A. (2015). Collocation in beginner learner writing: A longitudinal study. *System*, 53, 148-160.

Siyanova-Chanturia, A., & Pellicer-Sánchez, A. (2020). *Understanding formulaic language: A Second language acquisition perspective.* Routledge.

Snoder, P. (2017). Improving English Learners' productive collocation knowledge: The effects of involvement load, spacing, and intentionality. *TESL Canada Journal, 34*(3), 140-164.

Sonbul, S. (2015). Fatal mistake, awful mistake, or extreme mistake? Frequency effects on off-line/on-line collocational processing. *Bilingualism: Language and Cognition*, *18*(3), 419-437.

Sonbul, S., & Schmitt, N. (2013). Explicit and implicit lexical knowledge: Acquisition of collocations under different input conditions. *Language Learning, 63*(1), 121-159.

Sonbul S., & El-Dakhs, D. (2020). Timed versus untimed recognition of L2 collocations: Does estimated proficiency modulate congruency effects? Applied Psycholinguistics, 41(5), 1197-1222.

Storm, B. C., Bjork, R. A., & Storm, J. C. (2010). Optimizing retrieval as a learning event: When and why expanding retrieval practice enhances long-term retention. *Memory & Cognition*, 38, 244-253.

Szudarski, P. (2012). Effects of meaning- and form-focused instruction on the acquisition of verb-noun collocations in L2 English. *Journal of Second Language Teaching and Research*, 1, 3-37.

Szudarski, P. (2017). Learning and teaching L2 collocations: Insights from research. *TESL Canada Journal*, 34, 205-216.

Suzuki, Y., Nakata, T., & DeKeyser, R. (2020). Empirical feasibility of the Desirable Difficulty Framework: Toward more systematic research on L2 practice for broader pedagogical implications. *The Modern Language Journal, 104*(1), 313-319.

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

Toppino, T. C., & Gerbier, E. (2014). About practice: Repetition, spacing, and abstraction. In B. H. Ross (Ed.), *Psychology of learning and motivation* (Vol. 60, pp. 113-189). Academic Press.

Tsai, M.H. (2020). The effects of explicit instruction on L2 learners' acquisition of verb-noun collocations. *Language Teaching Research, 24*(2), 138162.

Uchihara, T., Webb, S., & Yanagisawa, A. (2019). The effects of repetition on incidental vocabulary learning: A meta-analysis of correlational studies. *Language Learning, 69*(3), 559-599.

Webb, S., & Nation, I. S. P. (2017). *How vocabulary is learned*. Oxford University Press.

Webb, S., Sasao, Y., & Balance, O. (2017). The updated Vocabulary Levels Test: Developing and validating two new forms of the VLT. *ITL – International Journal of Applied Linguistics*, 168(1), 34-70.

Wolter, B., & Gyllstad, H. (2011). Collocational links in the L2 mental lexicon and the influence of L1 intralexical knowledge. *Applied Linguistics, 32*(4), 430-449.

Wolter, B., & Gyllstad, H. (2013). Frequency of input and L2 collocational processing: A comparison of congruent and incongruent collocations. *Studies in Second Language Acquisition*, *35*(3), 451-482.

Yamashita, J., & Jiang, N. A. N. (2010). L1 influence on the acquisition of L2 collocations: Japanese ESL users and EFL learners acquiring English collocations. *TESOL Quarterly, 44*(4), 647–668.

Table 1

Number of Correct Responses and Percentages under All Conditions (N = 83)

Condition	Form Recall				Form Recognition			
	Pretest		Posttest		Pretest		Posttest	
	Correct	%	Correct	%	Correct	%	Correct	%
Equal spacing (*n* = 27)[a]	46	7.4	470	75.7	274	44.1	592	95.3
Expanding spacing (*n* = 25)[b]	53	9.2	419	72.9	240	41.7	516	89.7
Control (*n* = 31)[c]	89	12.5	142	19.9	370	51.9	427	59.9

[a] Max score = 621 (27 x 23)

[b] Max score = 575 (25 x 23)

[c] Max score = 713 (31 x 23)

Table 2

The Best-Fit Model for Variables Predicting Form Recall Scores (N = 1909, Log-Likelihood = −758.80, marginal R^2 =0.43 , conditional R^2 = 0.57)

	β	*SE*	*z value*	*p*	*Exp (β)*[a]	*Cohen's d*
(Intercept)	−2.50	0.28	-8.88	<.001***	0.08	-1.38
W1 length	−0.44	0.12	-3.72	<.001***	0.64	-0.24
Log vocabulary score	0.97	0.16	6.04	<.001***	2.64	0.53
Pretest score	3.26	0.33	10.00	<.001***	25.98	1.80
Condition: Equal spacing	4.30	0.39	10.92	<.001***	73.98	2.37
Condition: Expanding spacing	3.92	0.39	10.09	<.001***	50.39	2.16

Random effects:	*Variance*	*SD*
Participant	1.38	1.17
Item	0.21	0.46

[a] Odds ratio calculated based on log odds estimates (*β*)

*** $p < .001$

Table 3

The Best-Fit Model for Variables Predicting Form Recognition Scores (N =1909, Log-Likelihood = −638.20, marginal R^2 = 0.31, conditional R^2 = 0.40)

	β	*SE*	*z value*	*p*	*Exp (β)*[a]	*Cohen's d*
(Intercept)	−0.02	0.20	−0.09	.93	0.98	-0.01
W2 length	−0.26	0.11	−2.40	.02*	0.77	-0.14
Log vocabulary score	0.65	0.14	4.58	<.001***	1.92	0.36
Pretest score	1.18	0.17	6.90	<.001***	3.25	0.65
Condition: Equal spacing	3.74	0.42	8.89	<.001***	41.91	2.06
Condition: Expanding spacing	2.60	0.32	8.07	<.001***	13.51	1.44
Log vocabulary score x Condition: Equal spacing	0.75	0.43	1.73	.08	2.12	0.41
Log vocabulary score x Condition: Expanding spacing	0.76	0.36	2.11	.04*	2.14	0.42

Random effects:	*Variance*	*SD*
Participant	0.53	0.73
Item	0.14	0.38

[a] Odds ratio calculated based on log odds estimates (*β*)

* *p* < .05

*** *p* < .001

Figure 1

A Graphical Presentation of the Spacing Schedules Under the Equal and Expanding Spacing Conditions (REP = Repetition)

EQUAL SPACING 7-7-7																					
Treat-ment							**REP 1**							**REP 2**							**REP 3**
DAY-> **0**	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Treat-ment	**REP 1**					**REP 2**															**REP 3**
EXPANDING SPACING 1-5-15																					

Figure 2

A Graphical Presentation of the Interaction Between Condition and Log Vocabulary Score for the Form Recognition Measure. Shaded Areas Show 95 Per Cent Confidence Intervals (CI)