System 103 (2021) 102649

Contents lists available at ScienceDirect

System

journal homepage: www.elsevier.com/locate/system

The effect of spacing on incidental and deliberate learning of L2 collocations

Marijana Macis[a,*], Suhad Sonbul[b], Rezan Alharbi[c]

[a] Manchester Metropolitan University, UK
[b] Umm Al-Qura University, Saudi Arabia
[c] Kind Saud University, Saudi Arabia

ARTICLE INFO

Keywords:
Spacing
L2 collocations
Incidental learning
Deliberate learning
Form recall

1. Introduction

The effect of spaced practice has long been the focus of research in the field of cognitive psychology (e.g., Cepeda et al., 2006; Karpicke & Bauernschmidt, 2011; Logan & Balota, 2008). This enormous body of the literature has shown robust advantages of *distributed* over *massed practice*, which means that there will be more retention when the material to be learned is repeated on several separate occasions rather than when all the repetitions are presented in a single session (Cepeda et al., 2006). Spacing effects have also been found in various domains of second language learning (e.g., Bird, 2010; Li & DeKeyser, 2019; Rogers, 2015, 2017; Serrano & Muñoz, 2007; Suzuki, 2017) including vocabulary learning (e.g., Koval, 2019; Nakata, 2015; Nakata & Elgort, 2020; Nakata & Suzuki, 2019; Nakata & Webb, 2016; Rogers & Cheung, 2018). However, these studies have mainly examined the acquisition of single words. The effect of spacing on larger units of meaning, such as collocations, has been far less researched (Farvardin, 2019; Snoder, 2017). Given the importance of collocations for receptive and productive fluency and given the challenge they pose for second language learners (Boers, 2020), it is surprising how few L2 vocabulary studies have explored whether the spacing effect also holds for these and other multiword items. Thus, in order to shed light on this issue, the present study has two objectives, a) to explore how different

* Corresponding author. Department of Languages, Information and Communications, Faculty of Arts & Humanities, 5.12 Arts and Humanities Building, Cavendish Street, Manchester, M15 6BG.

E-mail address: m.macis@mmu.ac.uk (M. Macis).

https://doi.org/10.1016/j.system.2021.102649
Received 18 November 2020; Received in revised form 13 September 2021; Accepted 14 September 2021
Available online 21 September 2021

encounters with the same collocation should be spaced in order to maximise learning and b) to establish how far apart repetitions have to be in contexts where the focus of the activity is either on meaning or form. Addressing these aims would allow us to evaluate the potential of different spacing techniques for teaching and learning L2 collocations at the productive level of mastery.

2. Literature review

2.1. Effects of spacing on single words

Recently, there has been an increased interest in the effects of spacing on both intentional and incidental learning of vocabulary,[1] primarily single words (Elgort & Warren, 2014; Koval, 2019; Nakata, 2015; Nakata & Elgort, 2020; Nakata & Suzuki, 2019; Rogers & Cheung, 2018). For example, Nakata and Suzuki (2019) explored the effect of spacing on the intentional learning of semantically related (vs unrelated) word pairs. Target words were presented four times to EFL Japanese students, either "massed" (where a target word occurs several times in a short span of text) or "spaced" (where the occurrences a target word are widely scattered throughout in a text). Posttests (translation from L1 into L2) revealed that spacing had a positive effect for both types of word pairs. Similar findings have been reported in other spacing studies that investigated deliberate vocabulary learning (e.g., Karpicke & Bauernschmidt, 2011; Koval, 2019; Nakata, 2015).

Several theoretical interpretations have been proposed to account for the facilitative role of spaced practice in incidental learning. According to study-phase retrieval/reminding accounts (Toppino & Bloom, 2002), each spaced encounter with the target word elicits retrieval, which strengthens a memory trace formed by the previous encounter. But in massed exposure, the preceding memory trace is still active, which means that it cannot be retrieved, hence there are fewer retrieval opportunities and less learning. Encoding variability accounts (Benjamin & Tullis, 2010) hold that when a learner repeatedly encounters an item, and does so in different contexts, these contextual differences result in a greater number of retrieval routes, which facilitates long-term retention. Massed encounters, on the other hand, do not offer as much contextual variation and therefore, generate fewer retrieval cues.

In contrast to research on intentional vocabulary learning, research examining different spacing schedules in incidental learning contexts has yielded mixed findings. In a study by Serrano and Huang (2018), Taiwanese EFL students read and listened to a text seeded with target words, doing so either once every week for five consecutive weeks (spaced condition) or once every day for five consecutive days (massed condition). It was found that spacing was better than massing, however there was no statistically significant difference between the two conditions as measured by a delayed bilingual matching posttest. Different results were reported by Elgort and Warren (2014) and Rodgers and Webb (2019). Elgort and Warren measured knowledge through meaning generation and lexical decision tasks, whereas Rodgers and Webb assessed learning gains via form-meaning recognition tests. In contrast to Serrano and Huang, the studies found significantly higher learning gains when students encountered target words within the same chapter/episode (massed condition) as opposed to encountering them across different chapters/episodes (spaced condition). The differences between these two studies and that of Serrano and Huang may be due to a difference in the operationalisation of massing. In other words, what might be considered massed distribution in one study (e.g., daily encounters with a word over five days in Serrano and Huang) may be interpreted as spaced exposure in another (e.g., some students in Elgort and Warren took more than one day to read a chapter). In yet another relevant study on incidental learning, Nakata and Elgort (2020) found that spaced exposure was more effective for the acquisition of explicit knowledge of vocabulary, measured with meaning-recall and meaning-form matching tests. In the massed condition, Japanese EFL learners read three sentences containing the target word simultaneously. In the spaced condition, each of the three sentences was presented individually and separated by forty-seven intervening sentences. As pointed out by the authors, one explanation for the statistically significant gains in the spaced condition of this study might be the increased provision of feedback this group received.

Taken together, the reviewed studies suggest that spacing has a facilitative effect in the contexts of intentional learning of vocabulary. However, in situations where learners acquire vocabulary incidentally, the picture seems to be less clear. One important point to note is that these studies were conducted on single words only. Research suggests that vocabulary is also built from multiword items (e.g., idioms, collocations, phrasal verbs), which are crucial for successful language use (e.g., Crossley et al., 2015; Schmitt, 2010), but are difficult for L2 learners (e.g., Foster, 2001; Laufer & Waldman, 2011; Nesselhauf, 2003; Nguyen & Webb, 2017). Therefore, the questions that are worth asking are whether distributed learning practice might also facilitate the learning of multiword sequences, more specifically collocations, and to what extent the type of learning situation makes a difference.

2.2. Collocations

With increased research on L2 vocabulary acquisition during the last three decades, researchers have become particularly interested in the nature and learning of collocations, especially those that do not have a direct equivalent in L1, referred to as *incongruent collocations* (Nguyen & Webb, 2017; Wolter & Gyllstad, 2011; Yamashita & Jiang, 2010). Collocations are defined in this study, following the frequency-based approach (e.g., Sinclair, 1991), as combinations of two words that frequently co-occur together, regardless of their degree of semantic transparency. One of the limitations of this approach is that using measures of probabilistic strength of association as the only criterion leads to the classification of idioms as collocations (e.g., Macis & Schmitt, 2017; Nguyen & Webb, 2017; Webb et al., 2013). However, this approach is ecologically valid because learners are likely to encounter collocations of varying degrees of transparency in the input.

Research has shown that success in collocational learning depends on a number of factors such as frequency of encounters, previous vocabulary knowledge, congruency, transparency, and degree of attention (for an exhaustive overview, see Boers, 2020). Of these factors, frequency in particular, plays an important role in situations of both incidental and intentional learning (Macis, 2018; Pellicer-Sánchez, 2017; Peters, 2014; Szudarski & Carter, 2016; Webb et al., 2013). Several important conclusions have been drawn from such studies. The first is that both conditions are conducive to learning, however, when there is a deliberate focus on unfamiliar collocations, retention rates can be substantially enhanced. Further, at least five encounters might be necessary for learning collocations in both learning contexts. Finally, through repeatedly encountering items in input, L2 learners can acquire both the meaning and form of new collocations.

However, while important, frequency is not the only variable affecting the learning process; and for the input to become intake, learners also need to notice unfamiliar vocabulary (Schmidt, 2001). The more learners pay attention to morphological, orthographic, prosodic, semantic and/or pragmatic features of an unknown item, the more likely it is that the new lexical aspects will be retained (e.g., Laufer, 2006; Schmidt, 2010). Deliberate learning situations, to some extent, create conditions for noticing (for example, through typographic enhancement); however; there are usually fewer opportunities for noticing during content-based activities. In fact, noticing is particularly important in the latter contexts for two reasons. Many collocations are composed of delexicalised verbs (e.g., *make a mistake*) or highly familiar words (e.g., *tall building*), and because of familiarity with the component words, learners are less likely to notice their partnerships with surrounding words in the input (Godfroid et al., 2013; Laufer, 2011; Pellicer-Sánchez, 2020; Wray, 2002). Moreover, when compared to single words, collocations occur much less frequently on an individual basis (Boers et al., 2014). This suggests that the exposure that second language learners receive can be rather limited and the time that elapses between two encounters may be too long for long-lasting memory traces to form (Boers, 2020; Durrant & Schmitt, 2010). Thus, it might be speculated that when learners encounter new collocations incidentally, these need to be closer to each other in order to be noticed and "picked up" successfully. In contrast, when learners acquire collocations deliberately, the intervals between the repetitions can be longer because explicit attention is relatively likely to be paid to each item every time it is encountered.

2.3. Spacing and collocations

Research on how spacing affects acquisition of collocations is extremely limited. Snoder (2017) investigated the effectiveness of spaced practice (among other constructs) for the learning of verb-noun collocations. Spacing was operationalised as an *expanding* learning schedule, where the three treatment sessions were conducted at gradually increasing intervals: Day 1, Day 7, and Day 16. In the *intensive* learning schedule, the three treatments were close in time to each other as much as logistically possible (i.e., Day 1, Day 2, and Day 4). Fifty-nine L2 Swedish learners of English performed a series of tasks (both form- and meaning-focused) and their knowledge was assessed using a form-recall (cued L1→L2 translation) test. Even though the expanding spaced group performed better (as indicated by the mean test scores), the potentially low sample size, as noted by Snoder, might have led to statistically insignificant differences between the two groups. In a similar study, Farvardin (2019) assessed the effect that uniform spaced retrieval (USR), expanded spaced retrieval (ESR) and massed retrieval (MR) had on the intentional learning of *lexical collocations* (e.g., *piece of advice*) by sixty-two Iranian learners of English. The USR group received the treatment spread across four sessions in a 2-2-2 interval, the ESR group had the same number of sessions but in a 1-2-3 interval, and the MR group was taught the target collocations all within one session. Unlike Snoder (2017), results of recognition (multiple-choice) and recall (translation from L1 into L2) posttests showed that ESR and USR groups did significantly better than the MR group.

The above review provides initial, however very limited evidence, that spaced encounters lead to greater gains in L2 collocational knowledge under deliberate learning conditions. The findings from incidental learning contexts are less clear. Therefore, further research is needed to find out how different encounters with collocations should be spaced under different learning conditions in order to enhance long-term retention of form-meaning mappings. The present study aims to fill this gap by addressing the following research questions:

1 Which experimental condition (spaced or massed), if any, leads to long-term gains in collocation knowledge at the form-recall level under incidental (Experiment 1) and deliberate (Experiment 2) learning conditions?
2 Which condition is the most effective in developing long-term collocation knowledge at the form-recall level under incidental (Experiment 1) and deliberate (Experiment 2) learning conditions?

We hypothesised that the spacing effect may be present in the deliberate learning of collocations, but that massing may be necessary when collocations are acquired incidentally, as suggested by initial findings from research on single words (e.g., Rodgers & Webb, 2020). In fact, one might argue that collocations, especially those composed of familiar words, are less noticeable in context than unknown single words and might thus need massing even more than single words (e.g., Boers, 2020).

3. Experiment one

3.1. Method

3.1.1. Participants

The participants were 55 Arabic learners of English following an undergraduate degree in Translation (all females, age: 20–24 years, $M = 21.29$, $SD = 0.69$). On average they started learning English at 11.58 of age ($SD = 3.94$). Participants' average score on the Updated Vocabulary Levels Test (UVLT), Version A (Webb et al., 2017) at the 1K-word level was 29.31 out of 30 ($SD = 1.10$), indicating that they had a mastery of the 1000 most frequent words in English. They were also familiar with the 2K level, scoring 26.18/30 ($SD =$

3.26).

Based on logistical considerations (students' availability), the participants were pseudo-randomly assigned to the levels of Condition as follows: "Incidental Spaced (n = 18)", "(Incidental Massed (n = 19)", "Control (n = 18)". We did not experimentally control for the differences across the groups in vocabulary knowledge, however, the total VLT score out of 150 (five levels X 30 items) was later included as a covariate in the mixed-effects logistic regression analysis to control for any initial differences in estimated proficiency among participants in the various groups (e.g., Jaeger, 2008).

3.1.2. Items

The initial target pool included 39 potential incongruent adjective + noun collocations. However, after piloting the form recall test with natives (see *Measures* below), some items were removed and we ended up with 25 highly frequent adjective + noun collocations. Davies (2008a, 2008b) BNC (British National Corpus) and COCA (Corpus of Contemporary American English) interfaces were consulted to find adjectives that collocate with the most frequent 2000 nouns in English (Nation, 2012). The collocations had a COCA frequency of at least 5, a BNC frequency of at least 10, and a Mutual Information (MI) score of 3+ in both corpora (Webb & Nguyen, 2017).

Incongruency was controlled for, as it has been shown to affect collocation learning (Peters, 2016). To operationalise incongruency, three proficient Arabic speakers of English translated the individual constituent words from English to Arabic. Then, the second and third authors (both proficient Arabic-English speakers) translated the target collocations from English to Arabic to render a natural collocation in Arabic. Only collocations whose Arabic translation included the literal translation of one constituent word but not the other (e.g., *dead silence* = *pure silence* in Arabic) were considered incongruent and included in study.

We also controlled for collocation transparency as another major factor in collocation processing (see Gyllstad & Wolter, 2016). Ten native speakers of English were asked to rate the target 25 collocations for transparency (1 = opaque, 7 = transparent). The ratings were found reliable: Intraclass Correlation Coefficient *(ICC)* $(2,k)^2$ = .91, *95% confidence interval (CI)* [0.85, 0.96]. The mean rating score was 4.67 (*SD* = 1.41). Mean transparency ratings were included in the mixed-logit regression analysis as a covariate.

The target collocations were then divided into five sets of five collocations each to represent the five treatment weeks under the spaced and massed conditions (Appendix A). The following item-related variables were included as covariates in the mixed-logit regression analysis: collocation frequency, frequency of individual component words, and length of single words.

3.1.3. Design

There were two experimental conditions which represented two spacing schedules: *incidental spaced* and *incidental massed*. In the incidental learning condition, the target collocations were presented in a story five times each and were unenhanced. Five repetitions were included because previous research has indicated that a minimum of five repetitions seems to be necessary to learn a collocation (Peters, 2014; Webb et al., 2013). At the end of the text, there were five comprehension questions which measured the participants' comprehension of the story. These questions did not include the target items. In order to keep the total number of exposures under both conditions for each target collocation the same, these five repetitions were either distributed across five different sessions under the spaced conditions or collapsed all in one session under the massed conditions (see Table 1).

Since the massed conditions involved exposure to a different set of items in each session, we fit a separate mixed-logit regression model on a subset of data (massed condition only, see *Scoring and data analysis* below) prior to the main analysis where Item Set was tested as a covariate.

It should be noted that spacing/massing were operationalised in this way (with different number of items per session) to reflect spacing in an ecologically valid way. In previous research (e.g., Nakata & Suzuki, 2019), spacing and massing were operationalised as lexical items appearing one after another in a list (massed condition), or separated by other items in the same session (spaced condition). Nakata and Suzuki's (2019) operationalisation represents higher levels of control in an experimental setting whereby the number of target items to be learned in a single session is the same across conditions. We believe, however, that this does not reflect how spacing is usually done in language classrooms where it is common practice to revise vocabulary across learning sessions rather than in the same session. Since this study is intended to arrive at tangible implications for language teachers, we opted to operationalise spacing/massing across sessions, controlling for the number of total exposures. We do acknowledge though, that such a design might lead to uncontrolled differences across conditions due to differences in the number of items learned per session. This could in turn affect the level of engagement with the target items and the size of reported gains.

Table 1
Summary of the study design.

Session	Incidental spaced (stories)	Incidental massed (stories)
1	25 items x 1	5 items (Set 1) x 5
2	25 items x 1	5 items (Set 2) x 5
3	25 items x 1	5 items (Set 3) x 5
4	25 items x 1	5 items (Set 4) x 5
5	25 items x 1	5 items (Set 5) x 5
Total	**25 items x 5**	**25 items x 5**

3.1.4. Materials

The development of materials started with creating five different stories to represent the two incidental conditions. Each story was initially developed to include one encounter with each of the 25 target collocations (incidental spaced condition). Then, for the incidental massed condition, another version of the story was created to include five repetitions of five items in accordance with one of the five item sets (incidental massed condition). Lexical coverage was controlled for in all passages with 95% of the single words (excluding proper nouns) belonging to the most frequent 1000 word families in English. According to Nation (2006), 95% is the minimum lexical coverage percentage to allow for adequate unassisted reading comprehension. The stories were between 691 and 697 words long.

A set of five content-related comprehension questions was developed for each story and these were identical for both versions of each story. The five stories were piloted with native speakers of English in order to ensure naturalness. The final versions were then presented to three native speakers along with the comprehension questions. In the role of language learner, they all scored 100%. The average post-treatment comprehension score (across the five sessions) of the EFL learners in the study was also high (4.60 out of 5.00, *Min* = 4.00, *Max* = 4.80, *SD* = 0.25) showing excellent understanding.

3.1.5. Measures

For pretesting and posttesting we created two versions of a form recall test (Appendix B). The version differed only in the ordering of the test items. We opted for a cued form-recall test as it represents the real difficulty for non-natives learners who struggle with the production of collocations (e.g., Biskup, 1992). The test contained sentences with a defining context adapted from the BNC or COCA. Each sentence contained the noun node with the adjective removed. The cues included the number of spaces and the first one or two letter(s) of the target adjective. The number of letters included as a cue for each adjective was a result of several piloting rounds with native speakers. The target collocations were indicated in bold (*When President was ready to make his speech, there was a* ***d _ _ _ silence*** *in the room.)*.

The final piloting round involved ten native speakers of English who did not participate in any of the previous pilots. Two raters marked their responses as either correct (matching the target collocate) or incorrect (a blank or a plausible synonym) reaching 100% agreement. We selected the 25 collocations (out of 39) because at least eight out of the total ten natives provided the correct target adjective collocate (*M* = 95.2%, Min = 80%, Max = 100%, *SD* = 6.53). We consider the 80% threshold acceptable for the current gap-fill test as the remaining 20% mostly included low-frequency synonym collocates (e.g., “bland expression” instead of the more common “blank expression”). In addition, to ensure maximum comprehension of the contexts, any lexical item that was unlikely to be known by our participants was translated into Arabic in the margin. The reliability of the test was calculated based on pretest scores (Cronbach alpha = .88) and posttest scores (Cronbach alpha = .90) and was found to be acceptable.

3.1.6. Procedure

The whole treatment lasted for ten weeks (Appendix C). In Week 1, the Updated VLT and the pretest were administered. After a two-week gap, from Week 3 to Week 7, all participants in the experimental groups met with the researcher (the third author) outside the classroom once a week for 15 min each time. The sessions for each group took place on a different day of the week. However, a seven-day interval between sessions was maintained throughout the treatment period.

During each incidental learning session, the participants were instructed to read a story for 10 min and pay attention to the content. When they finished reading, the sheet containing the story was collected to avoid any extra exposure to target items. The participants were then asked to answer comprehension questions (5 min). The delayed posttest was administered in Week 10. Participants in the control groups did not receive any treatment and they continued with their usual translation classes. However, they took the pretest and the delayed posttest and also completed a language background survey along with all other participants.

3.1.7. Scoring and data analysis

The gap-fill test was scored on a binary 0/1 basis by the third author. Correct answers received a score of 1 and those that were wrong or left blank were scored 0. As the focus was on the link between the collocates rather than productive knowledge of single words, an answer that was misspelled but recognisable was also scored 1 (e.g., “nitive” for “native”). To check the reliability of the scoring procedures, 20.0% of the test items were scored by a second rater. Inter-rater reliability score was high *(ICC (1,1))*[3] = 0.93*, CI [*0.92, 0.94], indicating reliable scoring.

The analysis was conducted in R version 4.0.5 (R Core Team, 2021) through a mixed-logit regression model in which the binary dependent variable was posttest score. The focal independent variable was Condition, with the three levels: incidental massed, incidental spaced, and control. Control was the reference level. Pretest scores were included as a covariate in the model to control for any potential differences in pre-knowledge across participants in the various conditions. Finally, several item-related variables (Word 1 (W1) length, Word 2 (W2) length, W1 log frequency, W2 log frequency, log collocation frequency, and transparency rating), and one subject-related variable (Updated VLT scores as a rough measure of proficiency) were included as covariates in the model. All frequencies were extracted from the BNC and were log transformed to reduce skewness. Random effect structures of the models included random intercepts for items and participants and random by-item slopes for Condition.

We started with the null model (dependent measure and random variables only) and added factors incrementally in the forward method. We started by adding the item-related and participant-related covariates to control for their effect. This was followed by pretest scores and finally the focal Condition factor first as a fixed factor and then as by-item slopes. Likelihood ratio (X^2) tests were used at each step to check whether the inclusion of additional predictors contributed significantly ($p < .05$) to the model. To check consistency, we also conducted the analysis in the backward method with all predictors included in the initial model[4] and then removed them stepwise to test for their significant contribution. The resulting best-fit models were identical in both directions ($p < .05$).

As indicated above, a preliminary mixed-logit regression model was fit on the results for the massed condition only to test for any recency effect.[5] The model included Pretest scores as a covariate. Results showed that Item Set (1 through 5) did not significantly improve the model fit ($X^2(4) = 3.70$, $p = .45$). A summary of the continuous variables is presented in Appendix D. The continuous variables in all models were checked for collinearity. There were no collinearity issues as all VIF (variance inflation factor) values were below 2.

3.2. Results

3.2.1. Raw gains

Table 2 below presents the number and percentage of correct/incorrect responses in the gap-fill test in both the pre- and posttests and under the three conditions. In the pretest, the percentage of correct responses ranged from 28.0% to 32.4% for the three conditions. In the posttest, however, results varied more. The control group did show a test-retest effect with an increase of 6.67 percentile points. As for the two experimental groups, their pre-post gains over the control group (i.e., after excluding the 6.67 percentile points for a potential test-retest effect) were as follows: the incidental massed condition with 17.96 percentile points (around four collocations out of 25) and then the incidental spaced condition with 6.22 percentile points (around one collocation out of 25).

3.2.2. Mixed-logit regression analysis

Table 3 presents the best-fit model for the significant variables predicting post-test recall scores in the gap-fill test. It shows that a 1-unit increase in rough proficiency (as estimated by Updated VLT scores) is associated with a 2.0% increase in the odds of a correct answer. The much larger odds ratio for 'Pretest score' indicates that target collocations known before treatment are likely to be recalled about 14 times more often than collocations not known before treatment.

Furthermore, the effect of our focal variable (i.e., Condition) shows that participants in the incidental massed condition are about four times more likely to recall a target collocation than participants in the control group. This medium effect size corresponds approximately to Cohen's $d = 0.73$. The incidental spaced condition led to a similar (though much smaller) effect with an increase of 61.0% in the odds of a correct posttest response over the control condition. Finally, in order to inspect the remaining massed/spaced comparison, the reference level was redefined to 'incidental massed', showing a significant difference ($\beta = -0.86$, $SE = 0.23$, *z value* $= -3.77$, $p < .001$, *odds ratio* $= 0.42$, $d = 0.47$); participants in the incidental spaced condition are about half as likely to correctly recall a collocation than participants in the incidental massed condition. Accordingly, it can be concluded that the effect of the levels of the incidental condition was graded (i.e., incidental massed > incidental spaced > control)). Fig. 1 graphically presents the effect of Condition on posttest scores after controlling for the pretest scores and other significant covariates in the model.

4. Experiment two

4.1. Method

4.1.1. Participants

The participants were 50 L1 Arabic, L2 English learners from the same Translation programme who did not take part in Experiment 1 (all females, age: 19–26, $M = 20.82$, $SD = 1.27$). On average they started learning English at 12.02 of age ($SD = 3.38$). Their average scores on 1-K and 2-K word levels in the Updated VLT (Webb et al., 2017) were as follows: 1K ($M = 29.68$ out of 30, $SD = 0.65$) and 2K ($M = 28.12$ out of 30, $SD = 2.49$).

Like in Experiment 1, the participants were pseudo-randomly divided into three groups in a between-subject design: "Deliberate Spaced (n = 17)", "Deliberate Massed (n = 17)", "Control (n = 16)". The total VLT score was added as a covariate in the mixed-logit regression analysis to control for any potential differences in estimated proficiency among participants prior to treatment.

Table 2

Responses and percentages under all conditions in both testing sessions (Experiment 1, N = 55).

Condition	Pretest				Posttest			
	Correct	%	Incorrect	%	Correct	%	Incorrect	%
Incidental spaced [a]	135	30.0%	315	70.0%	193	42.9%	257	57.1%
Incidental massed [b]	154	32.4%	321	67.6%	271	57.1%	204	43.0%
Control [a]	126	28.0%	324	72.0%	156	34.7%	294	65.3%

[a] Max score = 450 (n = 18 x k = 25).

[b] Max score = 475 (n = 19 x k = 25).

Table 3
Summary of the best fit mixed-logit regression model, Experiment 1 (N = 1375, log-likelihood = −656.90).

Fixed effects	*β*	*SE*	*z value*	*p*		Odds ratio *(≈d)*
(Intercept)	−1.46	0.23	−6.27	<.001	***	0.23 (−0.80)
Updated VLT Score	0.02	0.01	3.60	<.001	***	1.02 (0.01)
Pretest score	2.65	0.19	13.93	<.001	***	14.15 (1.46)
Condition: Incidental massed	1.33	0.23	5.69	<.001	***	3.78 (0.73)
Condition: Incidental spaced	0.47	0.24	2.01	.04	*	1.61 (0.26)
Random effects:	*Variance*					
Subject (Intercept)	0.21					
Item (Intercept)	0.51					

$^{*}p < .05$, $^{***}p < .001$.

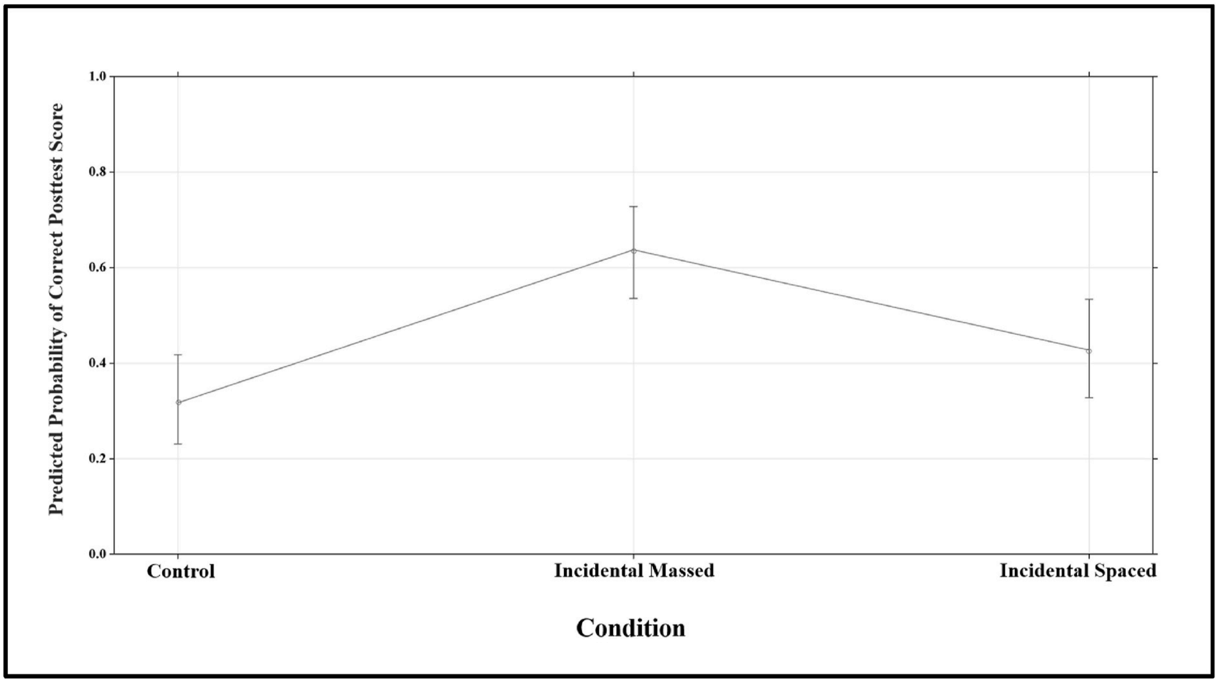

Fig. 1. Effects associated with the three levels of Condition (Experiment 1) when pretest scores are taken into account, with 95% error bars for the estimates.

4.1.2. Items

This experiment used the same 25 incongruent adjective + noun collocations included in Experiment 1 divided into the same five item sets.

4.1.3. Design

The same design employed in Experiment 1 was followed in Experiment 2. There were two treatment conditions and one control condition. The experimental conditions involved deliberate learning of collocations embedded in concordance lines in either massed or spaced fashion. Each target collocation was presented five times in five different concordance lines that were based on the stories developed for the incidental conditions. The concordance lines were followed by two form-focused exercises that focused the participants' attention on the target items.

4.1.4. Materials

Each target collocation was presented in a concordance line with five words on each side, and each collocation was further encountered in two recognition exercises, one matching and one multiple choice (see Supplementary Materials). The exercises were decontextualised and receptive to make them different from the pre-/posttest (a contextualised recall measure). It is important to note that the target items were not presented as wholes in these exercises to avoid additional exposure. The scores for both deliberate conditions averaged across sessions were high (4.74 out of 5.00 in the matching exercise, *Min* = 3.60, *Max* = 5.00, *SD* = 0.39; 4.99 out of 5.00 in the multiple-choice exercise, *Min* = 4.80, *Max* = 5.00, *SD* = 0.03), showing adequate practice.

4.1.5. Measures

The same recall measure developed in Experiment 1 was employed in this experiment. The reliability of the test was high based on pretest scores (Cronbach alpha = .84) and posttest scores (Cronbach alpha = .86).

4.1.6. Procedure

The procedures of Experiment 2 followed those of Experiment 1 in all aspects (see Appendix C). The only difference is that during the treatment, the participants were instructed to study the underlined target collocations in concordance lines and try to commit them to memory (7 min). They were informed that at the end of the task they were going to do exercises based on these collocations and that they would not be able to revisit the concordance lines. Therefore, after the study period, the sheets comprising the concordance lines were collected and the participants were instructed to do a matching exercise (4 min). Then, the sheet with this exercise was collected and a multiple-choice question (MCQ) exercise was administered (4 min). Thus, the total duration of each treatment session was the same (i.e., 15 min) across both experiments.

4.1.7. Scoring and data analysis

The scoring procedure was identical to Experiment 1. A second rater scored 20.0% of the test items with an inter-rater ICC *(1,1)* value of 0.95, CI *[*0.94, 0.96*],* indicating reliable scoring.

The analysis was conducted using mixed-logit regression models for binary data (posttest scores) in the same way detailed in Experiment 1. The fixed and random effects and slopes were the same and odds ratios were used as estimates of strength. The continuous item-related variables were the same as presented in Appendix D (no collinearity issues). The Updated VLT score ranged between 77 and 145 points *(centred* = −39.38 – 28.62*, SD* = 17.69*, Mdn* = 9.62*).* Also, like Experiment 1, a preliminary analysis was conducted on a subset of data (massed condition only) to test for the potential effect of Item Set on posttest scores. Results showed that adding Item Set did not significantly improve the model fit $(X^2(4) = 7.27, p = .12)$.[6]

4.2. Results

4.2.1. Raw gains

Table 4 shows that the pretest scores were low under the three conditions ranging from 27.8% to 32.0%, which suggests similar levels of knowledge before treatment. As for the posttest scores, the control group exhibited a test-retest effect with an increase of 7.75 percentile points. The two experimental conditions led to higher gains than the control group; the deliberate spaced condition with 37.90 percentile points (around nine collocations out of 25) followed by the deliberate massed with 19.54 percentile points (almost five collocations out of 25).

4.2.2. Mixed-logit regression analysis

The best-fit model is presented in Table 5 (same in both the forward and backward methods). Unlike Experiment 1, two item-related variables significantly predicted correct answers in the gap-fill test. The first was collocation frequency whereby a 1-unit increase in corpus occurrences was associated with a 43.0% increase in the odds of a correct posttest score. The second was transparency with 33.0% higher odds for a correct answer as the rating increased by one point.

The model also suggests that, similar to Experiment 1, a 1-unit increase in the Updated VLT scores is accompanied by a 2.0% increase in the odds of a correct posttest response. However, the effect of Pretest scores was much smaller in Experiment 2 than Experiment 1; items known in the pretest are over three times more likely to be recalled than collocations not known in the pretest.

Finally, the comparison of the three conditions in the posttest showed that they were significantly different (deliberate spaced > deliberate massed > control). Both the deliberate massed and deliberate massed conditions led to significantly higher gains than the control group with a large effect, but the deliberate spaced condition had a larger effect. Participants receiving this "deliberate spaced" treatment are 13 times more likely to recall the correct answer than those in the control group. The deliberate massed condition followed with around four times higher odds for a correct response than the control group. Finally, upon redefining the reference level to "deliberate massed", we find a significant effect $(\beta = 1.07, SE = 0.28, z\ value = 3.87, p < .001, odds\ ratio = 2.92, d = 0.59)$ that goes in the opposite direction of the effect reported in Experiment 1. Participants in the deliberate spaced condition are about two times more likely to correctly recall a collocation than participants in the deliberate massed condition. The effect of Condition on posttest scores is illustrated in Fig. 2.

Table 4
Responses and percentages under all conditions in both testing sessions (Experiment 2, N = 50).

Condition	Pretest				Posttest			
	Correct	%	Incorrect	%	Correct	%	Incorrect	%
Deliberate spaced [a]	135	31.8	290	68.2	329	77.4	96	22.6
Deliberate massed [a]	136	32.0	289	68.0	252	59.3	173	40.7
Control [b]	111	27.8	289	72.3	142	35.5	258	64.5

[a] Max score = 425 (n = 17 x k = 25).
[b] Max score = 400 (n = 16 x k = 25).

Table 5
Summary of the best fit mixed-logit regression model, Experiment 2 (N = 1250, log-likelihood = −662.8).

Fixed effects:	*β*	*SE*	*z value*	*p*		Odds ratio (≈*d*)
(Intercept)	−1.38	0.27	−5.20	<.001	***	0.25 (−0.76)
Log collocation frequency	0.36	0.14	2.51	.01	*	1.43 (0.01)
Transparency rating	0.29	0.11	2.53	.01	*	1.33 (0.16)
Updated VLT Score	0.02	0.01	3.34	<.001	***	1.02 (0.20)
Pretest score	1.20	0.19	6.20	<.001	***	3.32 (0.66)
Condition: Deliberate massed	1.51	0.31	4.79	<.001	***	4.51 (0.83)
Condition: Deliberate spaced	2.57	0.34	7.51	<.001	***	13.13 (1.42)
Random effects:	*Variance*					
Subject (Intercept)	0.17					
Item (Intercept)	0.87					
Condition (Deliberate massed)	1.04					
(Deliberate spaced)	1.29					

*p < .05, ***p < .001.

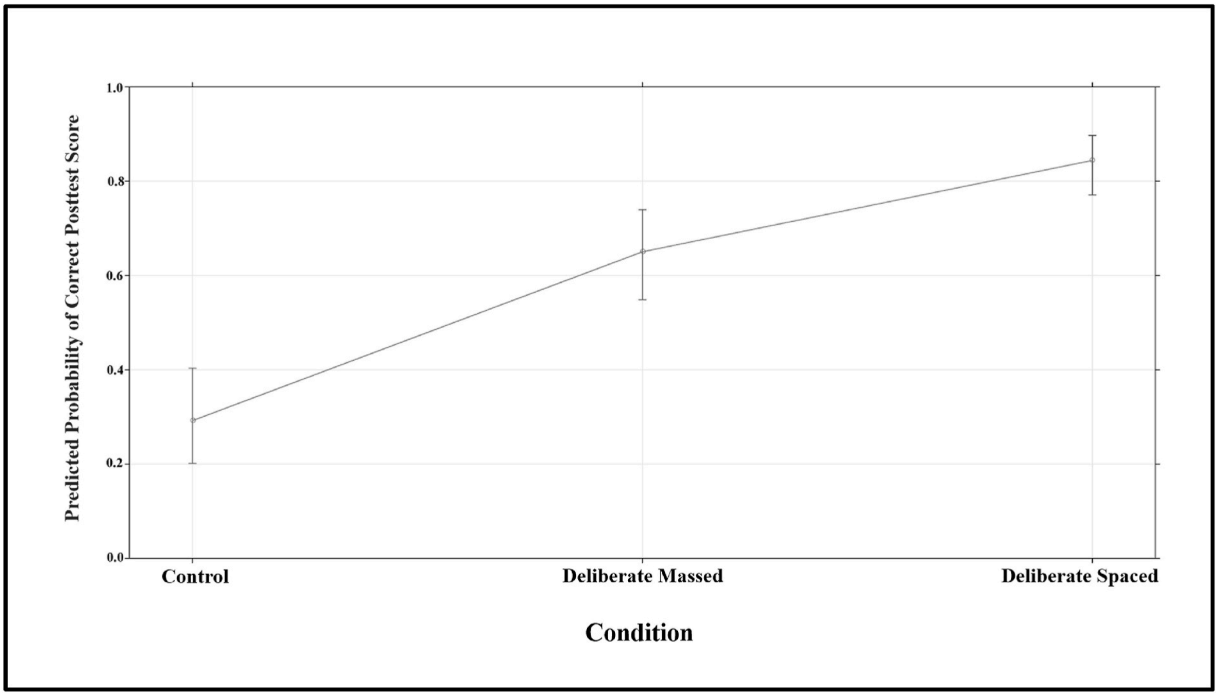

Fig. 2. Effects associated with the three levels of Condition (Experiment 2) when pretest scores are taken into account, with 95% error bars for the estimates.

5. Discussion

The present study investigated how different spacing schedules affected the long-term retention of collocations in incidental and deliberate learning conditions.

The first research questions asked which experimental condition (spaced or massed) would lead to long-term gains in collocation knowledge at the form-recall level under incidental and deliberate learning conditions. The results show that, when compared to the control group, all conditions led to vocabulary development (as measured on a three-week delayed posttest). Participants had mean vocabulary gains of four collocations out of 25 in the incidental massed condition and one collocation out of 25 in the incidental spaced condition (Experiment 1). In Experiment 2, participants scored an average of five collocations out 25 under massed schedule and nine collocations out 25 when the target collocations were spaced. Results are in line with previous research indicating that collocations can be acquired through both intentional and incidental approaches (for an overview, see Pellicer-Sánchez, 2020). These results also provide support for the advantage of form-focused instruction (FFI or instruction that is characterised by inducing L2 learners to pay attention to linguistic form) over meaning-focused activities in promoting EFL learners' collocational knowledge (Laufer, 2003; Nation, 2001; Sonbul & Schmitt, 2013). To account for this effectiveness, Laufer (2005, 2006) has related FFI to the Noticing hypothesis (Schmidt, 2001) according to which attention is a prerequisite for learning. In comparison to an incidental learning situation, a deliberate learning situation is one in which learners are encouraged to notice new vocabulary. In Experiment 2 of this study, the participants were explicitly instructed to pay attention to the target collocations and commit them to memory. It is possible that the deliberate nature of this learning situation created more opportunities for noticing, resulting in enhanced retention rates of collocational form knowledge.

Regarding our incidental learning condition, results suggest that both spacing schedules led to learning, however, the gains were significantly higher in the massed condition. One explanation is that some memory traces can be formed from spaced encounters in the input, however, they are limited; and to increase chances of learning, learners might need to have massed exposure (Boers, 2020). If encounters with target collocations are close to each other, it could be that the massing makes unknown collocations more salient and increases noticing, which, in turn, could increase learning gains. Moreover, while the mean number of collocations learned in the incidental massed group may seem small (four collocations learned out of 25), the posttest scores were still significantly higher than the incidental spaced condition and the control group, which suggests that massed exposure in written input has the potential to enhance collocation knowledge. Investigations into incidental learning of collocations, where all the repetitions of the target items were embedded within one text, have also reported similar amounts of learning (e.g., 2.18 collocations learned out of 18 in Webb et al., 2013). More importantly, if we look at the gains from a cost-benefit perspective (each session lasted only 15 min) and the level of knowledge we measured (productive, form-recall knowledge, that is particularly challenging for learners; see Schmitt, 2010), we would argue that learning collocations through massed exposure in texts is certainly worth the effort.

The second research question asked which condition was the most effective in developing long-term collocation knowledge at the form-recall level under incidental and deliberate learning conditions. The results indicate that the deliberate spaced condition was the most beneficial in developing long-term collocation knowledge. This group scored significantly higher than the deliberate massed group, which seems to suggest that distributed practice is more effective in deliberate learning situations. These findings support previous results on the spacing effect from cognitive psychology (e.g., Cepeda et al., 2006) as well as vocabulary studies conducted on intentional learning of single words (e.g., Nakata & Suzuki, 2019) and collocations (Farvardin, 2019). However, our results are contrary to Snoder's (2017) findings. This might be because the treatment in Snoder's study involved three constructs (involvement load, spacing and intentionality), so it is not clear whether the non-significant gains were due to spacing alone or the combined effect of the three variables. Moreover, no control group was included to control for the test-retest effect (as various measures were used), which might have also contributed to the lack of the spacing effect.

On the other hand, the results suggest that in incidental learning conditions, massing appears to be more beneficial. In fact, we would argue that massed exposure is especially important if we consider the relative scarcity of collocations in authentic input in comparison to single words (Boers & Webb, 2015) and the fact that some collocations are composed of high frequency words (as was the case in this study), which may also reduce the chances of the collocations being noticed (Pellicer-Sánchez, 2020). The findings of Experiment 1 fit with Elgort and Warren's (2014) and Rodgers and Webb's (2019) studies, which also indicate the advantage of massed exposure, when the focus is on single words. However, Elgort and Warren's results, in particular, need to be interpreted with caution as there was no delayed posttest, thus no claims can be made about long-term retention. Our findings go against Serrano and Huang (2018), who found no statistical difference between spaced and massed practice under incidental learning conditions. When interpreting the discrepancy between their findings and our study, we must take into account the methodological differences. In particular, the posttests were administered at different times for the two experimental groups (four days vs. 28 days) in Serrano and Huang as they were interested in the relationship between practice (spaced or massed) and an optimal testing time. Our study adopted the same delay (three weeks after treatment), which, in turn, may explain the disparity in results. Also, in Serrano and Huang multiple posttests have been administered, so the immediate posttest could have represented an opportunity for the retrieval of learned items, thus influencing the gains assessed by the delayed posttest. In our study, in order to address this concern and make stronger claims about long-term retention (e.g., Schmitt, 2010), we only administered one delayed posttest. The results of the present study are also different from Nakata and Elgort (2020) who found an advantage of spaced over massed presentation for single words in contextual word learning. One reason, as already stated, might be the provision of feedback, which may have benefitted the spaced group. No such feedback was provided in our study (Experiment 1). Furthermore, unlike our study, the target words in Nakata and Elgort (2020) were clearly identified (presented in brackets). Their participants were instructed to infer the meanings of the target words from context, and they could verify their inferences by checking target word definitions after reading. Yet, this study could still be considered incidental, since participants were not instructed to learn the words; instead, they were instructed to read for meaning. Thus, the difference in the effect of spacing versus massing may not necessarily be due to the incidental versus deliberate learning mode but it may be rather related to the opportunities the learning mode creates for noticing vocabulary targets in the input. This would explain why Elgort and Warren (2014), who did not use any typographic enhancement, found an advantage for words occurring within the same chapter, over words in different chapters, but Nakata and Elgort (2020) did not. It would also explain why, in the present study, massing worked better for incidental learning (as the target collocations were not highlighted or otherwise identified in the text), but spacing was more effective in the deliberate condition, in which collocations were underlined and shorter concordance lines were used. To sum up, when considering the spacing effects, we should take into account different learning conditions under which collocations are learned as this study provides initial evidence that they make a difference.

5.1. Effects of participant-related and item-related variables

The results of both experiments show that UVLT scores significantly predicted correct answers in the form-recall test. The finding that vocabulary size had an effect on collocational gains is, by no means, unexpected. Research has consistently shown that learners with a larger vocabulary size tend to learn more new lexical items (e.g., Peters, 2014). Moreover, transparency had a significant effect, but only on the deliberate learning of collocations. We speculate that once the learners' attention was explicitly directed to the target collocations, it was easier to learn those collocations that were more transparent (e.g., *sad fact*) than those that were more opaque in meaning (e.g., *white noise*). Transparency has indeed been shown to positively influence the likelihood of learning (Boers, 2020). Finally, the results of Experiment 2 show that corpus-derived frequency had an impact on learning outcomes. This might suggest that some of the target collocations may have been partially known and that highly frequent collocations might have made a greater impact on the deliberate than on the incidental learning conditions.

6. Pedagogical implications and limitations

The present study has important implications for both teachers and materials designers. Results highlight the importance of focusing L2 learners' attention on the target input through deliberate activities (e.g., Hulstijn, 2003). This suggests that including deliberate learning as part of any vocabulary programme will yield better results than one that does not include this important strand (Webb & Nation, 2013). Findings also indicate that a few exposures with target collocations (our study suggests five repetitions spaced weekly) may be enough for durable learning to occur if these encounters are spaced effectively. So, to improve learners' retention of collocations in intentional learning contexts, teachers should ensure that the target collocations are further practiced after they have been first introduced. This could be done both in the classroom (e.g., short quizzes) and through self-study (e.g., vocabulary learning apps such as *Anki*).

If the aim is to enhance collocational knowledge through reading, massed exposure appears to be more effective. This can be achieved through text manipulation as opportunities for encountering the same collocation multiple times in authentic input within a short span of time are scarce (Boers & Lindstromberg, 2009). However, to ensure that the items are embedded in a natural way in a text, resourcefulness and native-like knowledge of the items are required. Therefore, this implication might be more suitable for material designers who have the time and the required resources at their disposal (Boers, 2020).

Inevitably, this study suffers from a number of limitations. The spaced and massed conditions in our study differed in the number of items studied in every session (five versus 25 per session), not enabling us to make a direct comparison between the two conditions. We acknowledge that such a design might have affected the level of engagement with the target items and the size of reported gains. However, despite this built-in difference, it is worth pointing out that massed and spaced exposure had differential effects under the incidental and deliberate conditions. More research in this area is certainly needed.

It should also be noted that this study only measured the form-recall knowledge of collocations, so research focusing on other levels of collocational mastery, such as meaning, would be of great value as it would provide a clearer picture of the benefits of spacing and massing across various levels of collocational mastery. Finally, our study was not designed to measure noticing directly, thus future research should include a consideration of this variable. Despite these limitations, our study provides initial evidence that both massed and distributed practice might be relevant for the learning of collocations depending whether they need to be learned incidentally or deliberately.

Notes

1. The terms deliberate and intentional are used interchangeably in this study. We adopt Hulstijn's (2003) definition of intentional and incidental learning. Intentional learning is the learning that occurs when there is a deliberate intention to commit language features to memory, whereas incidental learning is defined as the learning that occurs when learners' attention is focused on understanding massages, rather than the form.
2. This ICC form stands for two-way random effects, absolute agreement, multiple raters' measurements (McGraw & Wong, 1996).
3. This ICC form stands for one-way random effects, absolute agreement, single rater/measurement (McGraw & Wong, 1996).
4. It should be noted that the initial full model did not converge. Thus, the model was refit with the 'bobyqa' optimiser which led to successful convergence (see Winter 2019 for more details).
5. This preliminary analysis tested for the same fixed effects, random intercepts, and slopes as the main analysis. The only fixed factor that was not tested in this preliminary analysis was Condition as only results from the massed condition were included. Similar to the main model (see Table 3), this preliminary model had Updated VLT scores and Pretest scores as main predictors. Additionally, the log frequency of W2 (word 2) was a significant predictor in the preliminary analysis.
6. Similar to the main model presented in Table 5, transparency rating, log collocation frequency, and Pretest scores were all significant predictors (but not the Updated VLT scores).

Author statement

Marijana Macis: Project administration, Conceptualisation, Supervision, Methodology, Visualisation, Writing - original draft, Writing - review & editing.
Suhad Sonbul: Project administration, Conceptualisation, Supervision, Methodology, Formal analysis, Writing - original draft.
Rezan Alharbi: Methodology, Investigation, Writing - original draft, Funding acquisition.

Declaration of competing interest

None.

Acknowledgements

We would like to thank Professor Norbert Schmitt for his useful input on the initial study design. We are also grateful to him and Eva Puimège for their insightful comments on an earlier version of this article. Any shortcomings are entirely our responsibility.

This research project was supported by a grant from the Research Centre for the Humanities, Deanship of Scientific Research, King Saud University.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.system.2021.102649.

Appendix A. Target collocations

No.	Collocation	Item Set	Transparency rating (1 = opaque, 7 = transparent)	BNC collocation frequency	BNC MI score	COCA collocation frequency	COCA MI score
1.	Soft drink	1	2.50	49	6.74	555	8.08
2.	Natural leader	1	6.10	19	3.9	94	3.79
3.	Short notice	1	5.20	253	7.15	475	6.79
4.	Native tongue	1	6.30	30	8.94	257	8.05
5.	Dead silence	1	4.10	14	4.43	180	5.31
6.	Blank expression	2	4.70	12	6.68	78	7.4
7.	Poor taste	2	4.10	20	5.08	125	5.17
8.	Little chat	2	5.60	40	5.65	91	4.42
9.	General public	2	5.90	721	5.6	2653	5.55
10.	Loving home	2	5.20	11	3.93	72	3.67
11.	Final leg	3	2.50	32	5.34	98	5.01
12.	Hard line	3	2.10	66	3.65	487	3.61
13.	Great pain	3	4.40	53	3.99	214	3.18
14.	Broken promises	3	6.40	16	7.13	255	8.48
15.	Painful death	3	6.40	11	4.87	75	4.85
16.	Sick joke	4	3.00	19	7.69	70	6.52
17.	Heavy smoker	4	4.30	27	10.61	56	8.81
18.	Low opinion	4	4.40	55	5.46	79	3.92
19.	Dark secret	4	4.10	26	5.21	146	4.66
20.	Strong interest	4	6.00	60	3.79	302	4.44
21.	Fine art	5	3.90	297	7.23	1313	6.61
22.	Small comfort	5	6.20	15	3.4	92	3.79
23.	Sad fact	5	4.90	52	5.41	148	4.38
24.	White noise	5	2.10	28	4.74	343	5.2
25.	Easy walk	5	6.30	19	3.7	81	3.09

Appendix B. Form recall test

Instructions

Example:

We have a **l _ _ _ _ amount** of money that we can spend. *(answer: large amount)*	spend: يصرف

Please try your best and answer all items. Also, please stick to the first letter(s) and the number of missing letters provided for each adjective.

Thanks very much for your participation!

التعليمات

يهدف الاختبار أدناه إلى تقييم معرفتك بـ 25 مترادفة لفظية (الكلمات التي عادة ما تظهر سوية) مؤلفة من صفات و أسماء، مثلاً 'large amount' . الكلمة الأولى من كل مترادفة تم حذفها. المطلوب منك هو تعبئة الفراغ بالصفة (كلمة تصف إنسان أو جماد) التي عادة ما تظهر مع الاسم المحدد في السياق المقدم. الشرطات تمثل عدد الحروف لكل صفة مطلوبة. لمساعدتك، تم توفير الحرف الأول (الحروف الأولى) من كل صفة والترجمة العربية للكلمات الموضوع تحتها خط.

مثال:

We have a **l _ _ _ _ amount** of money that we can spend. *(large amount* :الإجابة*)*	spend: يصرف

نرجو بذل مافي وسعك للإجابة على جميع الفقرات. أيضاً، يرجى الإلتزام بالحرف الأول (الحروف الأولى) و بعدد الشرطات المتوفرة لكل صفة.

شكراً جزيلاً لمشاركتك.

Sentence	Arabic Translation
1. If you want to learn how to create beautiful objects such as paintings, the University of London offers a degree in **f _ _ _ art**.	
2. They knew they were going to lose their jobs, so the manager's kind words were nothing more than a **sm _ _ _ comfort** to them.	
3. In the world cup next year, Pepsi is going to be the <u>official</u> **s _ _ _ drink**.	<u>official</u>: رسمي
4. Her husband showed no <u>emotion</u> and she was really angry at his **b _ _ _ _ expression**.	<u>emotion</u>: شعور
5. There are thousands of hungry children in the world but the <u>media</u> hides this very **s_ _ fact**.	<u>media</u>: وسائل الإعلام
6. I lied to my Mum that Dad had died and she was very angry and told me it was a **s _ _ _ joke**.	
7. The manager didn't need a lot of training to get the best out of his workers because he was a **n _ _ _ _ _ _ leader**.	
8. After taking two planes and two buses, he arrived in London and took a taxi for the **f _ _ _ _ leg** of his long journey home.	
9. By putting <u>criminals</u> in prison for longer times, the country had decided to take a **h _ _ _ line**.	<u>criminals</u>: مجرمين
10. The continuous sound of a machine like a hair dryer that often sends a crying baby to sleep is called **w _ _ _ _ noise**.	
11. There were only a few people in the park because they decided to arrange the <u>protest</u> at **s _ _ _ _ notice**.	<u>protest</u>: احتجاج
12. After my visit to the <u>dentist</u>, I was in such **g _ _ _ _ pain** that I couldn't even speak.	<u>dentist</u>: طبيب أسنان
13. The people are tired of all the government's **b _ _ _ _ _ promises**.	

. (*continued*).

Sentence	Arabic Translation
14. It was obvious from the yellow colour of her fingers that she was a **h _ _ _ _ smoker**.	obvious: واضح
15. Galicia is a region of Spain where people speak Galician as their **n _ _ _ _ _ tongue**.	region: منطقة
16. She, like many other girls who wrote to me, mentioned that the main cause of her depression was her **l _ _ opinion** of herself.	depression: **اكتئاب**
17. She was suffering from depression but hid this **da _ _ secret** from everyone.	suffering from:يعاني من depression: **اكتئاب**
18. We walked out in the middle of the second act because we found the play to be in **p _ _ _ taste**.	
19. I asked her to stay behind after class for a few minutes for a **l _ _ _ _ _ chat**.	
20. All the beaches in California are accessible to the **g _ _ _ _ _ _ public**.	accessible to: يمكن الوصول اليها
21. He suffered from cancer for many years which at the end led to a very slow and **pa _ _ _ _ _ death**.	suffered from: عانى من cancer: سرطان
22. All I want for my children is to grow up in a safe, **l _ _ _ _ _ home**.	
23. Despite having science degrees, he always had a particularly **st_ _ _ _ interest** in art and design.	despite: بالرغم من
24. When the President was ready to make his speech, there was a **d _ _ _ silence** in the room.	President: الرئيس speech**:** خطاب/ خطبة
25. We don't need a car to go the gym. It is just five minutes **e _ _ _ walk** from here.	gym: نادي رياضي

. (*continued*).

Appendix C. Experiment 1 procedure

Time	Condition	Tasks	Duration
Week 1	Incidental spaced	Updated VLT and the pretest	60 min
	Incidental massed		
	Control group		
Weeks 1–2	Gap		
Weeks 3–7 (five sessions with 7-day interval[a])	Incidental spaced	Reading short stories (10 min) + MCQ comprehension questions (5 min)	15 min
	Incidental massed		
	Control group	No treatment	
Weeks 8–10	Gap		
End (Week 10)		Language background survey and the posttest	40 min

a Participants who missed any treatment session were given an option to come the next day in order not to lose their data.

Appendix D. Summary of continuous variables (Experiment 1)

Variable	Range (adjusted range)	SD	Mdn
Log collocation frequency	2.40–6.58 (−1.19 – 3.00 log units)	1.03	−0.25
Log Word 1 frequency	2.64–6.43 (−1.96 – 1.70 log units)	1.00	0.26
Log W2 frequency	0.61–6.21 (−3.59–2.01 log units)	1.22	−0.08
W1 length	3.00–7.00 (−1.96 – 2.04 characters)	1.15	0.04
W2 length	3.00–10.00 (−2.48 – 4.52 characters)	1.68	−0.48
Transparency rating	2.10–6.40 (−2.57 – 1.73)	1.38	0.03
Updated VLT score	74.00–147.00 (−42.38 – 30.62 points)	17.69	1.62

Note. The second column shows the range of the variables. The adjusted range after centering, is presented in parentheses. Standard deviations and medians refer to the predictor values in the models. All variables are centred, and their means are zero.

References

Benjamin, A. S., & Tullis, J. (2010). What makes distributed practice effective? *Cognitive Psychology, 61*(3), 228–247. https://doi.org/10.1016/j.cogpsych.2010.05.004

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

Biskup, D. (1992). L1 influence on learners' renderings of English collocations: A polish/German empirical study. In P. J. L. Arnaud, & H. Bejoint (Eds.), *Vocabulary and applied linguistics* (pp. 85–93). Palgrave Macmillan.

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

Boers, F., & Lindstromberg, S. (2009). *Optimising a lexical approach to instructed second language acquisition*. Palgrave Macmillan.

Boers, F., Lindstromberg, S., & Eyckmans, J. (2014). Some explanations for the slow acquisition of L2 collocations. *Vigo International Journal of Applied Linguistics, 11* (1), 41–61.

Boers, F., & Webb, S. (2015). Gauging the semantic transparency of idioms: Do natives and learners see eye to eye? In R. R. Heredia, & A. B. Cieślicka (Eds.), *Bilingual figurative language processing* (pp. 368–392). Cambridge University Press.

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

Crossley, S. A., Salsbury, T., & McNamara, D. S. (2015). Assessing lexical proficiency using analytic ratings: A case for collocation accuracy. *Applied Linguistics, 36*(5), 570–590. https://doi.org/10.1093/applin/amt056

Davies, M. (2008a). The British national corpus (BNC): 560 million words, 1990-present. https://corpus.byu.edu/bnc/.

Davies, M. (2008b). The corpus of contemporary American English (COCA): 560 million words, 1990-present. https://corpus.byu.edu/coca/.

Durrant, P., & Schmitt, N. (2010). Adult learners' retention of collocations from exposure. *Second Language Research, 26*(2), 163–188. https://doi.org/10.1177/0267658309349431

Elgort, I., & Warren, E. (2014). L2 Vocabulary learning from reading: Explicit and tacit lexical knowledge and the role of learner and item variables. *Language Learning, 64*(2), 365414. https://doi.org/10.1111/lang.12052

Farvardin, M. T. (2019). Effects of spacing techniques on EFL learners' recognition and production of lexical collocations. *Indonesian Journal of Applied Linguistics, 9*(2), 395–403. https://doi.org/10.17509/ijal.v9i2.20237

Foster, P. (2001). Rules & routines: A consideration of their role in the task-based language production of native and non-native speakers. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: second language learning, teaching and testing* (pp. 75–94). Longman.

Godfroid, A., Boers, F., & Housen, A. (2013). An eye for words: Gauging the role of attention in L2 vocabulary acquisition by means of eye-tracking. *Studies in Second Language Acquisition, 35*(3), 483–517. https://doi.org/10.1017/S0272263113000119

Gyllstad, H., & Wolter, B. (2016). Collocational processing in light of the phraseological continuum model: Does semantic transparency matter? *Language Learning, 66* (2), 296–323. https://doi.org/10.1111/lang.12143

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

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*(4), 434–446. https://doi.org/10.1016/j.jml.2007.11.007

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. https://doi.org/10.1037/a0023436

Koval, N. G. (2019). Testing the deficient processing account of the spacing effect in second language vocabulary learning: Evidence from eye tracking. *Applied PsychoLinguistics, 40*(5), 1103–1139. https://doi.org/10.1017/S0142716419000158
Laufer, B. (2003). Vocabulary acquisition in a second language: Do learners really acquire most vocabulary by reading? Some empirical evidence. *Canadian Modern Language Review, 59*(4), 567–587. https://doi.org/10.3138/cmlr.59.4.567
Laufer, B. (2005). Focus on form in second language vocabulary acquisition. In S. H. Foster-Cohen, M. P. Garcia-Mayo, & J. Cenoz (Eds.), *EUROSLA yearbook* (vol. 5, pp. 223–250). John Benjamins.
Laufer, B. (2006). Comparing focus on form and focus on formS in second language vocabulary learning. *Canadian Modern Language Review, 63*(1), 149–166. https://doi.org/10.3138/cmlr.63.1.149
Laufer, B. (2011). The contribution of dictionary use to the production and retention of collocations in a second language. *International Journal of Lexicography, 24*(1), 29–49. https://doi.org/10.1093/ijl/ecq039
Laufer, B., & Waldman, T. (2011). Verb-noun collocations in second-language writing: A corpus analysis of learners' English. *Language Learning, 61*(2), 647–672. https://doi.org/10.1111/j.1467-9922.2010.00621.x
Li, M., & DeKeyser, R. (2019). Distribution of practice effects in the acquisition and retention of L2 Mandarin tonal word production. *The Modern Language Journal, 103*(3), 607–628. https://doi.org/10.1111/modl.12580
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. https://doi.org/10.1080/13825580701322171
Macis, M. (2018). Incidental learning of duplex collocations from reading: Three case studies. *Reading in a Foreign Language, 30*(1), 48–75. http://hdl.handle.net/10125/66738.
Macis, M., & Schmitt, N. (2017). Not just "small potatoes": Knowledge of the idiomatic meanings of collocations. *Language Teaching Research, 21*(3), 321–340. https://doi.org/10.1177/1362168816645957
McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1*(1), 30. https://doi.org/10.1037/1082-989X.1.1.30
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
Nakata, T., & Elgort, I. (2020). Effects of spacing on contextual vocabulary learning: Spacing facilitates the acquisition of explicit, but not tacit, vocabulary knowledge. *Second Language Research, 37*(2), 1–28. https://doi.org/10.1177/0267658320927764
Nakata, T., & Suzuki, Y. (2019). Effects of massing and spacing on the learning of semantically related and unrelated words. *Studies in Second Language Acquisition, 41*(2), 287–311. https://doi.org/10.1017/S0272263118000219
Nakata, T., & Webb, S. (2016). Does studying vocabulary in smaller sets increase learning? The effects of part and whole learning on second language vocabulary acquisition. *Studies in Second Language Acquisition, 38*(3), 523–552. https://doi.org/10.1017/S0272263115000236
Nation, I. S. P. (2001). *Learning vocabulary in another language.* Cambridge University Press.
Nation, I. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review, 63*(1), 59–82. https://doi.org/10.3138/cmlr.63.1.59
Nation, I. S. P. (2012). The BNC/COCA word family lists. Document bundled with Range program with BNC/COCA lists. https://www.victoria.ac.nz/lals/about/staff/paul-nation#publications, 25.
Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics, 24*(2), 223–242. https://doi.org/10.1093/applin/24.2.223
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. https://doi.org/10.1177/1362168816639619
Pellicer-Sánchez, A. (2017). Learning L2 collocations incidentally from reading. *Language Teaching Research, 21*(3), 381–402. https://doi.org/10.1177/1362168815618428
Pellicer-Sánchez, A. (2020). Learning single words vs. multiword items. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 158–173). Routledge.
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. https://doi.org/10.1177/1362168813505384
Peters, E. (2016). The learning burden of collocations: The role of interlexical and intralexical factors. *Language Teaching Research, 20*(1), 113–138. https://doi.org/10.1177/1362168814568131
R Core Team. (2021). *R: a language and environment for statistical computing.* Vienna, Austria: R foundation for statistical computing. https://www.R-project.org.
Rodgers, M. P. H., & Webb, S. (2019). Incidental vocabulary learning through viewing television. *ITL - international Journal of Applied Linguistics, 171*(2), 191–220. https://doi.org/10.1075/itl.18034.rod
Rogers, J. (2015). Learning second language syntax under massed and distributed conditions. *Tesol Quarterly, 49*(4), 857–866. https://doi.org/10.1002/tesq.252
Rogers, J. (2017). The spacing effect and its relevance to second language acquisition. *Applied Linguistics, 38*(6), 906–911. https://doi.org/10.1093/applin/amw052
Rogers, J., & Cheung, A. (2018). Input spacing and the learning of L2 vocabulary in a classroom context. *Language Teaching Research, 24*(5), 616–641. https://doi.org/10.1177/1362168818805251
Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). Cambridge University Press.
Schmidt, R. (2010). Attention, awareness, and individual differences in language learning. In W. M. Chan, S. Chi, K. N. Cin, J. Istanto, M. Nagami, J. W. Sew, T. Suthiwan, & I. Walker (Eds.), *Proceedings of CLaSIC 2010, Singapore, December vols. 2–4* pp. 721–737). National University of Singapore, Centre for Language Studies.
Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual.* Palgrave Macmillan.
Serrano, R., & Huang, H. Y. (2018). Learning vocabulary through assisted repeated reading: How much time should there be between repetitions of the same text? *Tesol Quarterly, 52*(4), 971–994. https://doi.org/10.1002/tesq.445
Serrano, R., & Muñoz, C. (2007). Same hours, different time distribution: Any difference in EFL? *System, 35*(3), 305–321. https://doi.org/10.1016/j.system.2007.02.001
Sinclair, J. (1991). *Corpus, concordance, collocation: Describing English language.* Oxford University Press.
Snoder, P. (2017). Improving English learners' productive collocation knowledge: The effects of involvement load, spacing, and intentionality. *TESL Canada Journal, 34*(3), 140–164. https://doi.org/10.18806/tesl.v34i3.1277
Sonbul, S., & Schmitt, N. (2013). Explicit and implicit lexical knowledge: Acquisition of collocations under different input conditions. *Language Learning, 63*(1), 121–159. https://doi.org/10.1111/j.1467-9922.2012.00730.x
Suzuki, Y. (2017). The optimal distribution of practice for the acquisition of L2 morphology: A conceptual replication and extension. *Language Learning, 67*(3), 512–545. https://doi.org/10.1111/lang.12236
Szudarski, P., & Carter, R. (2016). The role of input flood and input enhancement in EFL learners' acquisition of collocations. *International Journal of Applied Linguistics, 26*(2), 245265. https://doi.org/10.1111/ijal.12092
Toppino, T. C., & Bloom, L. C. (2002). The spacing effect, free recall, and two-process theory: A closer look. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(3), 437–444. https://doi.org/10.1037/0278-7393.28.3.437
Webb, S., & Nation, I. S. P. (2013). Teaching vocabulary. In C. A. Chapelle (Ed.), *The encyclopaedia of applied linguistics* (pp. 844–853). Wiley-Blackwell.
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
Webb, S., Sasao, Y., & Ballance, 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. https://doi.org/10.1075/itl.168.1.02web
Winter, B. (2019). *Statistics for linguists: An introduction using R.* Routledge.

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. https://doi.org/10.1093/applin/amr011

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge University Press.

Yamashita, J., & Jiang, N. (2010). L1 influence on the acquisition of L2 collocations. *Tesol Quarterly, 44*(4), 647–668. https://doi.org/10.5054/tq.2010.235998