RESEARCH ARTICLE

Effects of distributed practice on the acquisition of verb-noun collocations

Satoshi Yamagata[1]* (ID), Tatsuya Nakata[2] and James Rogers[3]

[1]Kansai University Dai-Ichi Senior High School/Dai-Ichi Junior High School, Japan, and University of Birmingham, UK; [2]Rikkyo University, Japan; [3]Meijo University, Japan
*Corresponding author. E-mail: yamagata@kochu.kansai-u.ac.jp or SXY034@student.bham.ac.uk

Abstract

Given the importance of collocational knowledge for second language learning, how collocation learning can be facilitated is an important question. The present study examined the effects of three different practice schedules on collocation learning: node massed, collocation massed, and collocation spaced. In the node-massed schedule, three collocations for the same node verb were studied on the same day. In the collocation-massed schedule, three collocations for the same node verb were studied in different weeks. In the collocation-spaced schedule, participants encountered multiple collocations for the same node verb within a single day; at the same time, multiple collocations for the same node verb were repeated each week. To examine whether the knowledge of studied collocations could be transferred to unstudied collocations containing the same node, posttests included novel collocations that were not encountered during the treatment. Results suggested that the collocation-spaced schedule led to the largest gains for both studied and unstudied collocations.

Introduction

Collocations refer to frequently recurring word combinations consisting of two content words (e.g., verb + noun, adjective + noun; Ackermann & Chen, 2013; Shin, 2006). Collocations can be adjacent (e.g., *meet demand*, *make decisions*) or nonadjacent (e.g., *meet the demand*, *make important decisions*; Boers et al., 2014; Wood, 2020), and are often characterized by restricted co-occurrence (e.g., *make a decision* but not *do a decision*; Paquot & Granger, 2012). Research suggests that collocational competence plays a pivotal role in second language (L2) learning, helping learners attain a sufficient level of accuracy and fluency (González-Fernández & Schmitt, 2015; Siyanova & Schmitt, 2008). Despite the importance of collocational knowledge, research also suggests that development of collocational knowledge often lags behind that of single words (Boers et al., 2014; Laufer & Waldman, 2011; Szudarski, 2017). Considering the importance of collocational knowledge for L2 learning, research examining how

collocation learning can be facilitated is valuable from both theoretical and pedagogical perspectives. Research has suggested that L2 collocation learning may be affected by various factors, such as frequency of exposure (e.g., Pellicer-Sánchez, 2017; Szudarski & Carter, 2016; Webb et al., 2013), type of input (e.g., Peters, 2016; Toomer & Elgort, 2019; Webb & Chang, 2022; Webb & Kagimoto, 2009), and type of exercise and instruction (e.g., Boers, Demecheleer et al., 2014; Boers et al., 2017; Eyckmans et al., 2016).

Another factor that can potentially affect L2 collocation learning is temporal spacing. Research indicates that distributing practice over a long period often facilitates the learning of single words (e.g., Bahrick & Phelps, 1987; Kim & Webb, 2022; Nakata & Webb, 2016). The memory advantage of spacing over no spacing (pure massing) is referred to as the *spacing effect*, whereas the advantage of longer temporal spacing over shorter spacing is referred to as the *lag effect* (Cepeda et al., 2006). Despite the potential benefits of spacing for vocabulary learning, research examining its effects on collocation learning is still limited. The present study aims to fill a gap in existing research by investigating the effects of spacing on the learning of L2 collocations.

Literature review

Research suggests that distributing practice opportunities over longer periods facilitates L2 vocabulary learning. For instance, in Nakata and Webb (2016, Experiment 2), 78 Japanese learners studied 20 English–Japanese word pairs under short- or long-spaced conditions. In the short-spaced condition, a target word was repeated after approximately 30 seconds, whereas in the long-spaced condition, a target word was repeated after approximately 3 minutes. Learning was measured by productive and receptive posttests conducted immediately and 1 week after the treatment. Posttest results suggested that long spacing was more than twice as effective as short spacing. Similarly, in Bahrick and Phelps (1987), 35 participants studied 50 English-Spanish word pairs using one of three spacing intervals: same day, 1 day, and 30 days. Retention was measured approximately 8 years after the last treatment. Posttest results suggested that spacing of 30 days was more than twice as effective as same-day spacing. Although the findings of these studies are useful, most studies have examined the effects of spacing on the learning of single words; thus, little is known about whether the benefits of spacing also extend to collocation learning. Two recent studies (Macis et al., 2021; Snoder, 2017), however, have examined the effects of spacing on L2 collocation learning and constitute exceptions.

Macis et al. (2021) compared the effects of massing and spacing on learning 25 adjective-noun collocations in incidental (Experiment 1) and deliberate learning conditions (Experiment 2). Across two experiments, Arabic EFL learners were assigned to one of five groups: incidental massed, incidental spaced, deliberate massed, deliberate spaced, and control. Those in the two incidental groups read short stories containing target collocations and answered comprehension questions (Experiment 1). Participants in the two deliberate groups studied the same target collocations through concordance lines and then completed matching and multiple-choice exercises (Experiment 2). In both massed and spaced groups, a given target collocation was encountered five times over 5 weeks. In the massed groups, a given target collocation was encountered five times on the same day. In the spaced groups, a given target collocation was encountered only once a week, and five occurrences were distributed over 5 weeks. Learning was measured by a fill-in-the-blank posttest, where participants

were asked to supply an appropriate adjective that collocated with the noun (e.g., the adjective *dead* for the noun *silence*). The results showed that the deliberate-spaced group had the largest gains, followed by the deliberate-massed, incidental-massed, and incidental-spaced groups. The findings suggest that spacing facilitates learning of not only single words but also collocations, albeit only in intentional learning.

Snoder (2017) conducted another study that examined the effects of spacing on L2 collocation learning. In it, 59 Swedish learners of English studied 28 verb-noun collocations under an expanding or intensive condition. In the expanding group, the treatment was given on days 1, 7, and 16, whereas in the intensive group, the treatment took place on days 1, 2, and 4. Learning was measured by a posttest that required participants to provide a verb that collocated with the noun. Posttests results showed no statistically significant difference between the two groups. One possible explanation for inconsistent findings between Macis et al. (2021) and Snoder (2017) may be that whereas the former compared massing (where practice opportunities for target materials are concentrated into a single day) and spacing (where practice opportunities for target materials are distributed over multiple days) and examined the spacing effect, the latter compared effects of two spacing schedulers (i.e., relatively short vs. relatively long intervals) and examined the lag effect.

Although the findings of these studies are informative, one potential limitation is that they examined the learning of only one collocate per node word. For instance, Snoder (2017) investigated the learning of 28 verb-noun collocations, but there was only one collocation for each node verb (e.g., *carry a risk*, *entertain hope*, *score success*). This is unfortunate because collocational development is perhaps facilitated by exposure to multiple collocations containing the same node. For instance, if Japanese learners of English first encountered the word *break* in the collocation *break a window*, they may associate *break* with its first language (L1) translation *waru* (i.e., to destroy a physical object; lexical association stage in Jiang, 2004), and may hypothesize that it can collocate only with concrete nouns. Exposure to collocations such as *break a promise* or *break the record* may allow learners to reconceptualize their knowledge of the meaning potential of the word (*semantic restructuring*; Jiang, 2004), helping them to comprehend or produce novel collocations such as *break a rule*, *break one's heart*, or *break the news*. Considering that exposure to multiple collocations with the same node perhaps facilitates collocational development, it would be useful to examine how spacing of multiple collocations with the same node affects collocational knowledge.

Multiple collocations for a given node may be introduced in one of the three schedules: massed from the perspective of the same node (hereafter, node massed), massed from the perspective of individual collocations (hereafter, collocation massed), and spaced from the perspective of individual collocations (hereafter, collocation spaced). In the node-massed schedule, practice opportunities for multiple collocations for the node word (e.g., *draw a line*, *draw tears*, *draw a conclusion*) are concentrated into a single session, and they are never repeated in subsequent sessions. In the collocation-massed schedule, multiple collocations for the same node are introduced across multiple sessions. For instance, learners may encounter *draw a line* in Day 1, *draw tears* in Day 2, and *draw a conclusion* in Day 3. However, practice opportunities for individual collocations are concentrated into a single session (e.g., *draw a line* is studied only in Day 1). As such, the schedule is massed from the perspective of individual collocations. In the collocation-spaced schedule, learners encounter multiple collocations for the same node word within a single session, just as in the node-massed schedule. At the same time, multiple collocations for the same node are repeated across multiple sessions. For instance, learners may study *draw a line*, *draw tears*, and *draw a*

conclusion in all Days 1, 2, and 3. Although it may be useful to examine how these three schedules affect collocation learning, the existing spacing studies on collocation learning (Macis et al., 2021; Snoder, 2017) have failed to provide evidence regarding the relative effectiveness of them.

When examining the effects of spacing on learning of multiple collocations with the same node, it would also be useful to examine not only the extent to which learners acquired collocations that they were exposed to but also the extent to which learners can transfer knowledge of studied collocations to novel, unstudied collocations that contain the same node but were not previously encountered (hereafter, *unstudied* collocations refer to novel collocations that contain the same node as the collocations that learners were exposed to). This is because exposure to multiple collocations with the same node may help learners to comprehend or produce novel collocations with the same node. For instance, exposure to multiple collocations with *break* (e.g., *break a window*, *break a promise*, or *break the record*) may allow learners to transfer knowledge of these collocations to novel, unstudied collocations with the same node (e.g., *break a rule*, *break one's heart*, or *break the news*).

Considering that exposure to multiple collocations with the same node is instrumental for collocational development, it would be useful to examine how the distributions of multiple collocations with the same node affect collocational knowledge. Furthermore, because learners do not acquire all collocations as individual units but instead make generalizations about which words can co-occur through repeated encounters with multiple collocations, examining the effects of spacing on knowledge of both studied and unstudied collocations would be useful. However, existing studies examining the effects of spacing on L2 collocation learning (Macis et al., 2021; Snoder, 2017), as well as the majority of previous studies on L2 collocation learning in general (e.g., Boers et al., 2017; Eyckmans et al., 2016; Pellicer-Sánchez, 2017; Peters, 2016; Szudarski & Carter, 2016; Toomer & Elgort, 2019; Webb & Chang, 2022), have failed to investigate the extent to which learners can transfer knowledge of studied collocations to unstudied collocations.

Investigating effects of spacing on unstudied collocations is also useful because it allows researchers to examine whether benefits of temporal spacing apply not only to recall of previously presented materials but also to induction. Specifically, transferring knowledge of studied collocations to unstudied collocations typically involves induction because it requires learners to extract the core, underlying meaning of the node, based on multiple collocations of the same node (unless learners are explicitly taught the core meaning of the node). Most studies examining effects of spacing on single words (e.g., Bahrick & Phelps, 1987; Nakata & Webb, 2016), in contrast, have investigated recall of previously presented materials (i.e., learners are presented with L2 words together with their meanings, and asked to learn them), rather than induction. It should be noted that some cognitive psychologists argue that although spacing may be effective for recall, it may not necessarily facilitate inductive learning. As Kornell and Bjork (2008) state, it is possible that "spacing is the friend of recall, but the enemy of induction" (p. 585). This is because presenting multiple instances of a particular category or concept simultaneously (i.e., massing) may help learners identify underlying conceptual features. In contrast, when multiple instances of a particular concept are presented after long intervals (spacing) learners may have difficulty noticing underlying commonalities, thus making induction more difficult. Given that spacing may have differential effects on recall and induction, it is possible that the learning of collocations, especially those of unstudied collocations, may benefit from spacing to a lesser degree than the learning of single words.

The present study

The present study aims to fill this research gap by investigating effects of node-massed, collocation-massed, and collocation-spaced schedules on knowledge of studied and unstudied collocations. The treatment was conducted over 3 weeks. In the node-massed schedule, multiple collocations for the same node verb (*draw a conclusion*, *draw a line*, *draw tears*) were concentrated into a single day and were not repeated in subsequent days. In the collocation-massed schedule, multiple collocations for the same node verb were introduced across the 3 weeks. For instance, participants studied *draw a line* in Week 1, *draw tears* in Week 2, and *draw a conclusion* in Week 3. However, practice opportunities for individual collocations were concentrated into a single session (e.g., *draw a line* was studied only in Week 1). In the collocation-spaced group, participants studied multiple collocations for the same node verb within a single day, just like the node-massed schedule. At the same time, multiple collocations for the same node verb were repeated each week. For instance, participants studied three collocations for the node verb *draw* (*draw a conclusion*, *draw a line*, *draw tears*) in Weeks 1, 2, and 3 (see "Method" section for more details).

This study will answer the following research question: To what extent do node-massed, collocation-massed, and collocation-spaced schedules facilitate knowledge of studied and unstudied L2 collocations?

The following hypotheses were formulated for the mentioned research question:

Hypothesis 1: For the retention of studied collocations, the collocation-spaced schedule will be more effective than the node-massed and collocation-massed schedules.

Hypothesis 2: For the knowledge of unstudied collocations, the collocation-spaced schedule will be the most effective, the collocation-massed schedule will be the least, and the node-massed schedule will be between the two.

Hypothesis 1 predicts that the collocation-spaced schedule will be most effective for the retention of studied collocations. This is because whereas practice opportunities for individual collocations are distributed over 3 weeks in the collocation-spaced schedule, they are concentrated into a single session in the two massed schedules. Existing studies have produced inconsistent results regarding effects of spacing on collocation learning. Specifically, whereas Macis et al. (2021) found benefits of distributed practice for intentional learning, Snoder (2017) failed to do so. Hypothesis 1 predicts that the results of this study will be consistent with those of Macis et al. (2021). This is because, just like the study conducted by Macis and colleagues, the present study involves the comparison of spacing and massing and examines the spacing effect, whereas Snoder's study involved the comparison of two spacing schedulers (i.e., relatively short vs. relatively long intervals) and examined the lag effect.

Hypothesis 2 predicts that for knowledge of unstudied collocations, the collocation-spaced schedule will be the most effective, followed by the node-massed schedule. In the node-massed and collocation-spaced schedules, learners are exposed to multiple collocations for the same node (e.g., *draw a line*, *draw tears*, *draw a conclusion*) within a single session, which may allow learners to extract the core, underlying meaning of the node. This in turn may improve the ability to transfer knowledge of studied collocations to novel, unstudied collocations that contain the same node. In the collocation-massed schedule, in contrast, multiple collocations with the same node are introduced in different weeks (e.g., Week 1: *draw a line*, Week 2: *draw tears*, Week 3: *draw a conclusion*). This may make it difficult for learners to notice underlying commonalities

motivating the use of the node word, resulting in limited ability to transfer knowledge of studied collocations to unstudied collocations.

Hypothesis 2 also predicts the advantage of the collocation-spaced schedule over the node-massed schedule for knowledge of unstudied collocations. Existing studies comparing blocking (i.e., a schedule where only one concept or skill is practiced at a time) and interleaving (i.e., a schedule where multiple concepts or skills are practiced at once) suggest that blocking may be beneficial for finding commonalities among different exemplars of a particular concept or category (Carpenter & Mueller, 2013; Kang, 2016), which might predict the advantage of the node-massed schedule over the collocation-spaced schedule for unstudied collocations. This is because whereas the node-massed schedule (where exemplars from only one node are presented each day) is akin to blocking, the collocation-spaced schedule (where exemplars from multiple nodes are presented each day) is akin to interleaving. However, although encounters with a given node are concentrated into a single session in the node-massed schedule, they are distributed over 3 weeks in the collocation-spaced schedule. Because encounters with a given node distributed over a longer period in the collocation-spaced schedule may help consolidate learners' understanding of the meaning potential of the node, Hypothesis 2 predicts the advantage of the collocation-spaced schedule over the node-massed schedule for knowledge of unstudied collocations.

Method

Participants

The original pool of participants consisted of 96 first-year Japanese EFL high school students (15–16 years old). Six students who missed one or more of the pretest, treatment, or posttest sessions were excluded from analysis, resulting in 90 participants. All participants had learned English in a formal setting for at least 4 years. Prior to the experiment, they took the 1,000 to 5,000 frequency levels on the Updated Vocabulary Level Test (UVLT), Version B (Webb et al., 2017). The average scores are provided in Table 1. The participants came from three intact classes, each of which was randomly assigned to one of three groups: node massed ($n = 27$), collocation massed ($n = 31$), and collocation spaced ($n = 32$). Because a statistically significant difference was found among the three groups in their total scores on UVLT, $F(2, 87) = 8.37, p < .001, \eta^2 = .16$ (collocation spaced > node massed [$p < .001$]; collocation spaced > collocation massed [$p = .039$]; collocation massed = node massed [$p = .084$]), the UVLT score was used as a covariate in the analysis (see the following text). An *a priori* power analysis for a mixed within-between 3 × 2 ANOVA (three groups at two measurement points) showed that when the effect size was set to be medium ($f = .25$), a minimum of 64 participants would be necessary. As a result, the number of participants in the present study ($n = 90$) was deemed sufficient.

Materials

Fifty-four verb-noun collocations (e.g., *carry weight*, *draw a line*, *take advice*) were chosen as target items. All collocations were incongruent between the L1 (Japanese) and L2 (English), that is, the translation of the node verb in each collocation was different from its most common, prototypical translation equivalent (Conklin & Carrol, 2018; Gyllstad & Wolter, 2016; Szudarski, 2012). Initially, 144 collocations were identified as candidates for target items. Based on results of a norming test administered to

Table 1. Proportion of correct responses on the UVLT

	Node massed		Collocation massed		Collocation spaced	
	M [95% CI]	*SD*	*M* [95% CI]	*SD*	*M* [95% CI]	*SD*
1,000-word levels	81.5% [77.2%, 85.8%]	10.8%	87.1% [83.0%, 91.2%]	11.1%	90.5% [88.3%, 92.7%]	6.1%
2,000-word levels	52.0% [44.9%, 59.1%]	18.0%	59.9% [53.4%, 66.4%]	17.3%	70.7% [66.0%, 75.5%]	13.1%
3,000-word levels	27.7% [19.5%, 35.8%]	20.5%	37.0% [31.2%, 42.8%]	15.7%	41.6% [36.1%, 47.1%]	15.3%
4,000-word levels	25.4% [18.4%, 32.5%]	17.9%	30.9% [26.0%, 35.7%]	13.0%	36.7% [30.1%, 43.3%]	18.3%
5,000-word levels	15.3% [10.1%, 20.5%]	13.2%	22.2% [17.5%, 26.8%]	12.2%	24.8% [19.8%, 29.8%]	13.9%
Total	40.4% [35.0%, 45.8%]	13.6%	47.5% [43.6%, 51.4%]	10.4%	52.9% [49.1%, 56.6%]	10.4%

191 Japanese high school students who did not participate in the actual experiment, the 144 items were narrowed down to 74 (see Appendix S1 in the Online Supplementary Materials). To identify collocations that were unfamiliar, a pretest was carried out 3 weeks before the treatment with actual participants of the experiment.

Two types of tests were given as the pretest: collocation filling and verb filling. In the collocation-filling test, participants were presented with a short sentence where a target collocation was deleted and asked to supply the missing verb and noun. To clarify the meaning of the target collocation, a Japanese translation of the sentence was provided. To prevent participants from providing alternate, acceptable answers (e.g., *run a fever* instead of *have a fever*), the number of letters was provided as a hint. In addition, based on a similar procedure utilized in Nakata and Webb (2016), a letter from the word was sometimes provided as a hint when deemed necessary to help avoid alternative, acceptable answers. Participants were informed that they could provide multiple answers if they could think of more than one. An example of a collocation-filling item is as follows:

もし熱が出れば、できるだけ早く私に言ってくださいね。
If you (_ _ _) a/an (_ _ _ _ _), please tell me as soon as possible.
(Answer. *run, fever*)

Cloze sentences were created so that all words used in each sentence would be among the most frequent 4,000 word families of the COCA. As the results of the UVLT suggest (see "Participants" section), it is possible that some participants were not familiar with some words used in these sentences. However, because Japanese translations for all sentences were provided, potential use of unfamiliar words perhaps did not have major effects on the results of this study. Vocabulary load analysis also showed that the most frequent 1,000 word families alone cover 95.4% of running words used in the cloze sentences.

In the verb-filling test, the noun of the candidate collocation was given, and participants were required to fill in the missing verb. After completing the collocation-filling test, participants were asked not to return to any items on the collocation-filling test. Both pretests required the production, rather than comprehension, of target collocations. Because research suggests that production of collocations poses more of a challenge for L2 learners than comprehension (Gyllstad & Wolter, 2016; Henriksen, 2013; Laufer & Waldman, 2011), this study is also concerned with the development of productive knowledge of collocations. As a result, productive collocational knowledge was measured in both pretests and posttests. The pretest is provided in Appendix S2 in the Online Supplementary Materials. Based on the results of a pilot study involving 80 Japanese learners recruited from a different high school than the school where the main study was conducted, participants were given up to 40 minutes to complete the pretest. They were also instructed to put a circle around the number of the last question they solved if they were unable to complete the test within the time limit. None of the participants indicated that they were unable to complete the pretest. Because it was not possible to identify a sufficient number of target collocations based on the results of the pretest, an additional pretest with 12 novel collocations was administered 2 weeks before the treatment (see Appendix S3 in the Online Supplementary Materials). Based on results of the pretest and additional pretest, 54 target collocations, which consisted of nine node verbs and their six collocate nouns, were chosen (see Appendix S4 in the Online Supplementary Materials). Out of the 54 target collocations, 53 were chosen from the first pretest, and only one item (*cut a loss*) was chosen from the additional pretest.

The target collocations were divided into two sets of 27 items (nine node verbs and their three collocate nouns each). One set of items was assigned to *studied* items, whereas the other to *unstudied* items. Both studied and unstudied items were tested on the pretest and posttest. However, although studied items were presented and practiced during the treatment, unstudied items did not appear throughout the treatment. Unstudied items were included to examine effects of the treatment on the ability to transfer knowledge of studied collocations to unstudied, novel collocations that contain the same node. The two sets were created so that they were matched for variables such as the average pretest score, *t*-score (Hunston, 2002; Webb et al., 2013), frequency in the Corpus of Contemporary American English (COCA), and familiarity ratings of the nouns by Japanese learners (see Appendix S4 in the Online Supplementary Materials). Care was also taken to ensure that collocations with similar meanings (e.g., *cut class* and *cut school*) would not be included in the same set. Only collocations that were semantically motivated by the core meaning of the node verb were used as target collocations. This is because otherwise we would not be able to expect learners to transfer the knowledge of studied collocations to unstudied collocations. The relationship between the core meaning of the node verb and the meaning of each target collocation is provided in Appendix S4 in the Online Supplementary Materials.

Procedure

Treatment

Three weeks after the pretest and 2 weeks after the additional pretest, the treatment was conducted over 3 weeks. Three treatment sessions were given each week (on Monday, Wednesday, and Friday), resulting in nine sessions in total. Each session took approximately 5 to 10 minutes and was conducted during regular class hours. Different target collocations were introduced in each class, depending on the participants' group (i.e., node massed, collocation massed, or collocation spaced). Figure 1 presents target collocations introduced in each session in the three groups. In the node-massed group, participants learned three collocations containing the same node verb each day (e.g., *draw a conclusion*, *draw a line*, *draw tears*). In the collocation-massed group, multiple collocations for the same node verb were studied in different weeks. For instance, participants studied *draw a line* in Week 1, *draw tears* in Week 2, and *draw a conclusion* in Week 3. In the collocation-spaced group, participants studied multiple collocations for the same node verb (e.g., *draw a conclusion*, *draw a line*, *draw tears*) within a single day. At the same time, multiple collocations for the same node were encountered every week throughout the treatment. For instance, participants studied three collocations for the node verb *draw* (*draw a conclusion*, *draw a line*, *draw tears*) in each week of the treatment.

For each treatment session, materials were presented on a screen in front of the classroom using presentation software. The treatment session consisted of the following seven stages: (1) presentation of target collocations, (2) presentation of target collocations in context, (3) retrieval of target verbs, (4) translation of target collocations, (5) retrieval of target verbs in context, (6) retrieval of target collocations in context, and (7) a quiz. See Appendix S5 in the Online Supplementary Materials for further details of the stages. Three out of the seven stages (Stages 2, 5, and 6) involved a context sentence containing a target collocation. For a given collocation, the same context sentence was used for all three stages. This is because a study conducted by Durrant and Schmitt (2010) suggests that repeating the same context sentence three times may facilitate L2 collocational development more than using three different sentences.

	Node massed				Collocation massed		
	Wednesday	Friday	Monday		Wednesday	Friday	Monday
Week 1	run a fever	carry a product	take advice		carry a product	cut class	draw a line
	run a story	carry a tune	take aim		run a story	make way	meet a goal
	run a finger	carry weight	take notice		take advice	set a record	put a stop
Week 2	cut class	set a tone	make way		carry weight	cut a price	draw a tear
	cut a deal	set a date	make mention		run a finger	make sense	meet a challenge
	cut a price	set a record	make sense		take notice	set a date	put a space
Week 3	put a stop	draw a conclusion	meet a challenge		carry a tune	cut a deal	draw a conclusion
	put emphasis	draw a tear	meet a demand		run a fever	make mention	meet a demand
	put a space	draw a line	meet a goal		take aim	set a tone	put emphasis

	Collocation spaced								
	Wednesday			Friday			Monday		
Week 1	run a fever	carry a product	take advice	cut class	set a tone	make way	put a stop	draw a conclusion	meet a challenge
	run a story	carry a tune	take aim	cut a deal	set a date	make mention	put emphasis	draw a tear	meet a demand
	run a finger	carry weight	take notice	cut a price	set a record	make sense	put a space	draw a line	meet a goal
Week 2	run a fever	carry a product	take advice	cut class	set a tone	make way	put a stop	draw a conclusion	meet a challenge
	run a story	carry a tune	take aim	cut a deal	set a date	make mention	put emphasis	draw a tear	meet a demand
	run a finger	carry weight	take notice	cut a price	set a record	make sense	put a space	draw a line	meet a goal
Week 3	run a fever	carry a product	take advice	cut class	set a tone	make way	put a stop	draw a conclusion	meet a challenge
	run a story	carry a tune	take aim	cut a deal	set a date	make mention	put emphasis	draw a tear	meet a demand
	run a finger	carry weight	take notice	cut a price	set a record	make sense	put a space	draw a line	meet a goal

Figure 1. Target items introduced in each treatment.

For Stages (1)–(7), the target collocations were presented in a block of three (for the node-massed and collocation-massed groups) or nine items (for the collocation-spaced group) and repeated, instead of one collocation going through all seven stages one by one. For instance, as shown in Figure 1, in the node-massed group, for the first treatment session (Wednesday in Week 1), the following three collocations were introduced: *run a fever, run a story, run a finger.* At the beginning of the treatment, all three collocations were presented for Stage (1). After this, the three collocations were practiced in Stage (2). This was followed by the three collocations practiced in Stage (3), and so forth. To minimize order effect, the items appeared in a different order for each stage.

As shown in Figure 1, whereas three collocations were practiced each day in the two massed groups, in the collocation-spaced group, nine collocations were practiced each day. Please note, however, that when collapsed across all treatment sessions, the number of encounters was held constant for all three groups. For instance, in the two massed groups, participants completed all seven stages for the target collocation *run a story* in the first treatment session in Week 1 (see Appendix S5 in the Online Supplementary Materials). In contrast, for the target collocation *run a story*, participants in the collocation-spaced group completed Stages (1), (2), and (7) in Week 1, Stages (3) and (4) in Week 2, and Stages (5) and (6) in Week 3. Because each target collocation was practiced seven times throughout the treatment in all three conditions, the number of encounters was held constant for all three groups. Because the treatment was paced by the presentation software, time-on-task was also held constant, and the only difference was how the practice opportunities were distributed.

Posttests
Immediately after the last treatment session (Monday in Week 3; Figure 1), participants took the immediate posttest. It was different from the pretest in four respects. First, although the number of letters, and sometimes one letter from the word, was provided as a hint in the pretest (e.g., _ _ _ _ for *take*), no hint was provided on the posttest. Second, in the pretest, 98 items (74 in the pretest and 24 in the additional pretest) were tested in both the collocation-filling and verb-filling tests. In the posttest, only 27 studied collocations were tested in the collocation-filling test, and 54 target collocations (27 studied and 27 unstudied) were tested in the verb-filling test. Unstudied items were not tested in the collocation-filling test (which required learners to provide both the node verb and collocate noun) because we cannot expect any of the treatments to contribute to the learners' ability to successfully provide the correct collocate noun, which was not encountered during the treatment. Third, a randomized item order different from the pretest was used for the immediate posttest to minimize order effect. Fourth, because the posttest involved less items than the pretest, the time limit for the posttest (20 minutes) was shorter than that for the pretest (40 minutes). The time limit for the posttest was determined based on a pilot study with 80 Japanese learners recruited from a different high school than the school where the main study was conducted. Other than these, the immediate posttest was the same as the pretest (see the immediate posttest in Appendix S6 in the Online Supplementary Materials). Two weeks after the immediate posttest, a delayed posttest was administered without prior announcement. This was identical to the immediate posttest except for item order. Two types of posttests (collocation filling and verb filling) were used in the present study. This is because administering two posttests with different levels of sensitivity may provide a more comprehensive picture regarding the incremental nature of collocation learning (Peters, 2016; Szudarski & Carter, 2016).

Scoring and data analysis

Collocation-filling test

The knowledge of intact, studied collocations was measured by the collocation-filling test. If participants provided both the verb and noun successfully, it was scored as correct. Misspelled responses were scored as correct as long as they were recognizable (e.g., Snoder, 2017; Sonbul & Schmitt, 2013; Toomer & Elgort, 2019). To control for effects of prior knowledge, for each participant, items answered successfully on the pretest were treated as missing values and excluded from analysis (e.g., Nakata & Suzuki, 2019). This resulted in the exclusion of 1.3% of items on average per participant (node massed: 1.1%, collocation massed: 1.2%; collocation spaced: 1.5%). All analyses had α levels set at .05.

Responses were analyzed using a mixed-effect logistic regression model with the *lme4* package (version 1.1-27.1; Bates et al., 2015) in R (version 4.1.2; R Core Team, 2021). The response variables were discrete binary data (correct = 1, incorrect = 0). Treatment (node massed, collocation massed vs. collocation spaced) and Test_timing (immediate vs. delayed posttest) were included as fixed effects. To control for English proficiency effects, UVLT scores were included as a covariate in the model. Furthermore, to control for recency effect, lag to test (the number of days between the last occurrence of the item during the treatment and immediate posttest) was also included as a covariate. For instance, the last occurrence for the target item *run a fever* during the treatment was in the first treatment session (Wednesday in Week 1) in the node-massed group, whereas it was in the seventh treatment session (Wednesday in Week 3) in the collocation-spaced group (Figure 1). Therefore, lag to test was 19 days for the node-massed group, and 5 days for the collocation-spaced group. Because the difference between 17 and 19 days, for instance, may be larger than the difference between 0 and 2 days, lag to test was squared before it was entered into the model. To avoid multicollinearity and convergence issues, both UVLT scores and squares of lag to test were centered and standardized before they were entered into the model as s.UVLT_score and s.Lag_to_test, respectively.

The random effects were fitted using the maximum likelihood method, assuming random intercepts for participants and target collocations, and random slopes for target collocations toward the UVLT scores and lag to test. An interaction between Treatment and Test_timing was also entered into the model, assuming that the treatment's effect was different for the immediate and delayed posttests.

Verb-filling test

Knowledge of verbs in studied and unstudied collocations was measured by the verb-filling test. The verb-filling test was scored in the same way as the collocation-filling test. As in the collocation-filling test, items answered correctly on the pretest were treated as missing values for each participant. This resulted in the exclusion of 2.1% of items on average per participant (node massed: 2.8%, collocation massed: 1.8%; collocation spaced: 1.7%).

The model used for the verb-filling test was the same as the one used for the collocation-filling test, except that collocation type (Studied = 1 and Unstudied = 0) was included as new fixed and random effects after it was centered and standardized (s. Collocation_type). Two interactions (Test_timing × Treatment × s.Collocation_type; Test_timing × s.Collocation_type) were also included. To make the model converge, we added an additional optimizer, and pipelines to random effects, which allow the exclusion of correlated parameters between random variables. Because the

Table 2. Proportion of correct responses on the pretest

	Collocation-filling test			Verb-filling test		
	M [95% CI]	*SD*	Range	*M* [95% CI]	*SD*	Range
Node massed	0.6% [0.3%, 1.1%]	2.2%	0%–11.1%	2.8% [2.1%, 3.8%]	6.5%	0%–33.3%
Collocation massed	0.6% [0.3%, 1.1%]	1.7%	0%–7.4%	1.8% [1.3%, 2.6%]	3.1%	0%–13.0%
Collocation spaced	0.8% [0.4%, 1.3%]	1.6%	0%–7.4%	1.7% [1.2%, 2.4%]	3.0%	0%–13.0%

unstudied items did not appear in any of the treatment sessions, dummy coding was used for s.Lag_to_test of these items.

Results

Pretest

Table 2 shows results of the pretest scores. More detailed information about the pretest performance is provided in Appendix S7 in the Online Supplementary Materials. The differences in the pretest scores of the three groups were not statistically significant, producing negligible effects; collocation-filling pretest: $H(2) = 1.88$, $p = .391$, $r = .09$, verb-filling pretest: $H(2) = 0.14$, $p = .933$, $r = .01$.

Learning-phase performance

The proportion of correct responses on the quiz given at the end of each treatment session (Stage 7; see Appendix S5 in the Online Supplementary Materials) was 96.4% (95% CI = [93.6%, 99.2%]; $SD = 7.6$%) for the node-massed group, 94.4% (95% CI = [92.0%, 96.9%]; $SD = 6.5$%) for the collocation-massed group, and 75.8% (95% CI = [70.2%, 81.5%]; $SD = 15.7$%) for the collocation-spaced group. The difference was statistically significant, $H(2) = 46.34$, $p < .001$, and a large effect size ($r = .68$) was detected, according to Plonsky and Oswald (2014) benchmark. Post-hoc analysis with Bonferroni adjustment revealed a statistically significant difference between the collocation-spaced and node-massed groups ($z = -5.87$, $p < .001$, $r = .76$ [large effect]), as well as between the collocation-spaced and collocation-massed groups ($z = -5.48$, $p < .001$, $r = .70$ [large effect]). No statistically significant difference, however, was found between the two massed groups ($z = -1.91$, $p = .170$, $r = .25$ [medium effect]). The findings suggest that the two massed groups led to higher scores than the collocation-spaced group during the learning phase.

Collocation-filling test

The reliability of the collocation-filling test indexed by Cronbach alpha was .917 for the immediate posttest and .904 for the delayed posttest, showing sufficient reliability (Plonsky & Derrick, 2016). Results for the collocation-filling test are summarized in Tables 3 to 5, as well as in Figure 2. Table 4 shows fixed and random effects in the mixed-effect logistic regression model. The significant fixed effect of the collocation-spaced group suggests that when collapsed across the immediate and delayed posttests, the collocation-spaced group significantly outperformed the node-massed group. The odds ratio (OR) of 9.58 indicates that the odds of being able to answer correctly on the
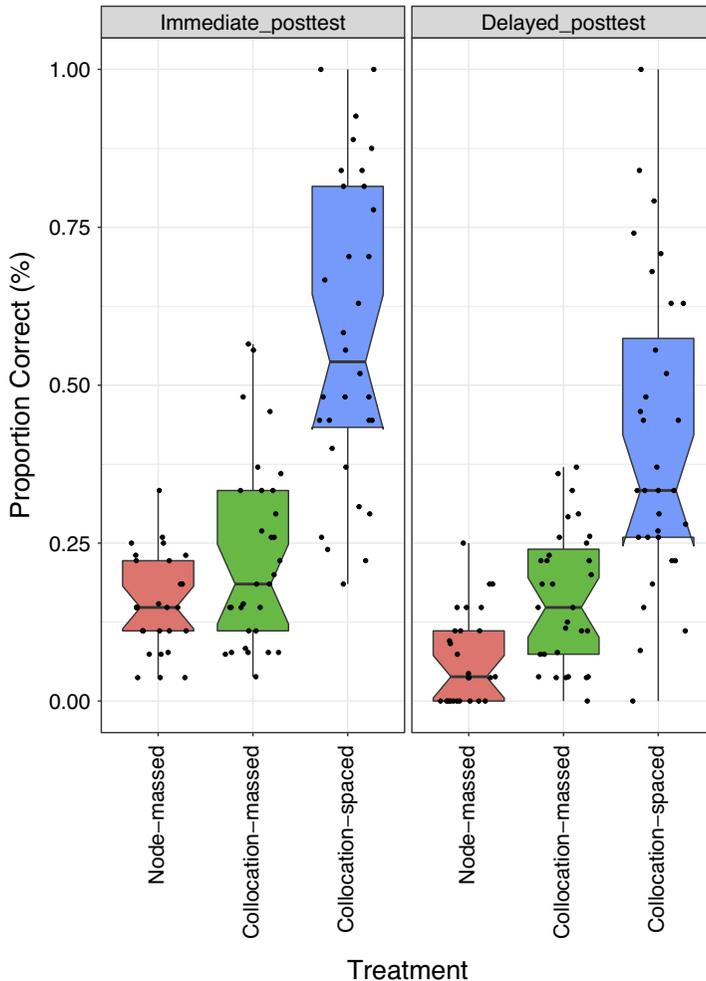
Figure 2. Distributions of scores for the collocation-filling test.

posttest in the collocation-spaced group were 9.58 times higher than the node-massed group, which is considered a large effect size, according to guidelines proposed by Chen et al. (2010), where odds of 1.68/3.47/6.71 are interpreted as small, medium, and large effects, respectively. The significant fixed effect of the UVLT suggests that higher UVLT scores were associated with higher posttest scores, with a small effect (OR = 2.36). The fixed effect of lag to test, however, was not statistically significant, which suggests that the recency effect (i.e., whether the last encounter with the target collocation was close to the posttest or not) did not significantly affect learning. Although the fixed effects of Collocation-massed and Test_timing were also significant, they are not discussed in detail because an interaction containing these effects was also significant (see the following text).

Table 3. Proportion of correct responses on the posttest

			Node massed		Collocation massed		Collocation spaced	
			M [95% CI]	*SD*	*M* [95% CI]	*SD*	*M* [95% CI]	*SD*
Immediate	Collocation filling		14.9% [12.2%, 17.6%]	35.6%	23.0% [20.1%, 25.9%]	42.1%	57.8% [54.5%, 61.2%]	49.4%
	Verb filling	Studied	16.4% [13.7%, 19.2%]	37.1%	28.2% [25.0%, 31.3%]	45.0%	62.1% [58.8%, 65.4%]	48.5%
		Unstudied	4.6% [3.1%, 6.2%]	21.0%	7.8% [5.9%, 9.6%]	26.8%	19.9% [17.2%, 22.5%]	39.9%
Delayed	Collocation filling		6.6% [4.7%, 8.5%]	24.8%	15.9% [13.4%, 18.5%]	36.6%	40.7% [37.4%, 44.1%]	49.2%
	Verb filling	Studied	9.4% [7.1%, 11.6%]	29.2%	20.2% [17.4%, 23.1%]	40.2%	45.6% [42.3%, 49.0%]	49.8%
		Unstudied	2.8% [1.5%, 4.1%]	16.6%	5.7% [4.1%, 7.3%]	23.2%	16.6% [14.1%, 19.1%]	37.2%

Table 4. List of fixed and random effect fitted: Collocation-filling test fixed effects

	Estimate	95% CI	SE	z	Odds Ratio	p
Intercept	−3.38	[−4.10, −2.65]	0.37	−9.09	0.03	<.001
Collocation spaced	2.26	[1.60, 2.91]	0.33	6.76	9.58	<.001
Collocation massed	1.09	[0.45, 1.73]	0.33	3.34	2.97	<.001
Test_timing	1.19	[0.77, 1.61]	0.21	5.57	3.29	<.001
s.UVLT_score	0.86	[0.60, 1.12]	0.13	6.48	2.36	<.001
s.Lag_to_test	−0.31	[−0.69, 0.06]	0.19	−1.65	0.73	.098
Collocation spaced × Test_timing	−0.21	[−0.69, 0.27]	0.24	−0.86	0.81	.388
Collocation massed × Test timing	−0.58	[−1.09, −0.07]	0.26	−2.25	0.56	.024

Model Formula: glmer (Phrase_accuracy ~ Treatment*Test_timing + s.UVLT_score + s.Lag_to_test + (1 | ID) + (s.UVLT_score + s.Lag_to_test + 1 | Item), Data, family = binomial, control = glmerControl (optimizer = "bobyqa"))

Random Effects

		Variance	SD
Participants	(Intercept)	0.65	0.81
Item	(Intercept)	1.71	1.31
	s.UVLT_score	0.11	0.34
	s.Lag_to_test	0.56	0.75

None of the interactions in the model were statistically significant except for the interaction between Collocation-massed × Test_timing (OR = 0.56). This significant interaction indicates that scores for the node-massed group decayed more than those for the collocation-massed group from the immediate posttest to the delayed posttest, widening the gap between the two groups. Post-hoc analysis with Tukey's test was conducted using the R package *lsmeans* (Lenth, 2021). Results (Table 5) showed that on the immediate posttest, the collocation-spaced group significantly outperformed both the node-massed (OR = 7.77 [large effect]) and collocation-massed groups (OR = 4.66 [medium effect]). The collocation-massed group, in contrast, failed to significantly outperform the node-massed group, producing a negligible effect (OR = 1.67). On the delayed posttest, the collocation-spaced group significantly outperformed both the

Table 5. Results of the post-hoc analysis for treatment: Collocation-filling test

Posttest	Comparisons	Estimate	95% CI	SE	z	Odds Ratio	p
Immediate	Collocation spaced vs. Node massed	2.05	[1.45, 2.64]	0.30	6.73	7.77	<.001
	Collocation spaced vs. Collocation massed	1.54	[1.02, 2.06]	0.27	5.78	4.66	<.001
	Collocation massed vs. Node massed	0.51	[−0.06, 1.07]	0.29	1.75	1.67	.500
Delayed	Collocation spaced vs. Node massed	2.26	[1.60, 2.91]	0.33	6.76	9.58	<.001
	Collocation spaced vs. Collocation massed	1.17	[0.64, 1.71]	0.27	4.29	3.22	<.001
	Collocation massed vs. Node massed	1.09	[0.45, 1.73]	0.33	3.34	2.97	.011

node-massed (OR = 9.58 [large effect]) and collocation-massed groups (OR = 3.22 [small effect]). Unlike on the immediate posttest, the collocation-massed group significantly outperformed the node-massed group, and a small effect was found (OR = 2.97). The findings suggest the following order on the collocation-filling test:

Immediate posttest: collocation-spaced > collocation-massed = node-massed
Delayed posttest: collocation-spaced > collocation-massed > node-massed

Verb-filling test

The reliability of the verb-filling test indexed by Cronbach alpha was .934 for the immediate posttest and .930 for the delayed posttest, showing sufficient reliability (Plonsky & Derrick, 2016). Results for the verb-filling test are summarized in Tables 3, 6, and 7, as well as Figure 3. Table 6 shows fixed and random effects in the mixed-effect logistic regression model. The significant fixed effect of the collocation-spaced group
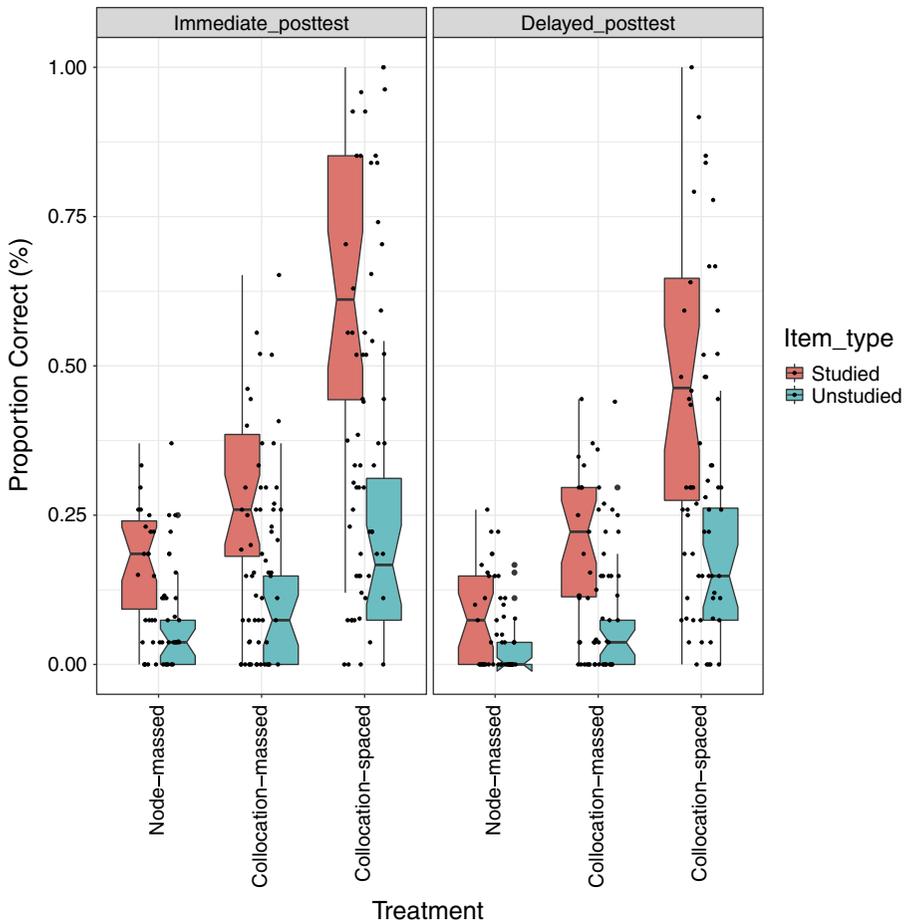

Figure 3. Distributions of scores for the verb-filling test.

Table 6. List of fixed and random effects fitted: Verb-filling test fixed effects

	Estimate	95% CI	SE	z	Odds Ratio	p
Intercept	−3.50	[−4.09, −2.90]	0.30	−11.48	0.03	<.001
Collocation spaced	1.70	[1.04, 2.36]	0.34	5.04	5.47	<.001
Collocation massed	0.49	[−0.16, 1.15]	0.33	1.48	1.63	.139
Test_timing	0.71	[0.36, 1.07]	0.18	3.92	2.03	<.001
s.Collocation_type	0.91	[0.44, 1.38]	0.24	3.81	2.48	<.001
s.UVLT_score	0.88	[0.62, 1.14]	0.13	6.60	2.41	<.001
s.Lag_to_test	−0.13	[−0.36, 0.09]	0.11	−1.18	0.88	.239
Collocation spaced × Test_timing	−0.04	[−0.45, 0.36]	0.21	−0.21	0.96	.834
Collocation massed × Test_timing	−0.20	[−0.64, 0.24]	0.22	−0.90	0.82	.370
Collocation spaced × s.Collocation_type	0.24	[−0.14, 0.63]	0.20	1.23	1.27	.220
Collocation massed × s.Collocation_type	0.28	[−0.13, 0.69]	0.21	1.35	1.32	.178
Test_timing × s.Collocation_type	0.11	[−0.24, 0.47]	0.18	0.63	1.12	.527
Collocation spaced × Test_timing × s.Collocation_type	0.23	[−0.17, 0.63]	0.21	1.12	1.26	.263
Collocation massed × Test_timing × s.Collocation_type	−0.03	[−0.47, 0.41]	0.22	−0.12	1.03	.903

Model Formula: glmer (Verb_accuracy ~ Treatment*Test_timing*s.Collocation_type + s.UVLT_score + s.Lag_to_test + (s. Collocation_type + 1 || ID) + (s.UVLT_score + s.Lag_to_test + s.Collocation_type + 1 || Item), Data, family = binomial, control = glmerControl (optimizer = "bobyqa", optCtrl=list(maxfun=2e5))

Random Effects

		Variance	SD
Participants	(Intercept)	0.95	0.98
	s.Collocation_type	0.12	0.34
Item	(Intercept)	0.82	0.91
	s.UVLT_score	0.08	0.28
	s.Lag_to_test	0.14	0.37
	s.Collocation_type	0.46	0.67

suggests that when collapsed across the immediate and delayed posttests and studied and unstudied collocations, the collocation-spaced group significantly outperformed the node-massed group, producing a medium effect (OR = 5.47). The fixed effect of the collocation-massed group, however, was not statistically significant, and only a negligible effect was observed (OR = 1.63). This suggests that when collapsed across the immediate and delayed posttests and studied and unstudied collocations, no significant difference existed between the two massed groups. The significant fixed effect of Test_timing shows that when collapsed across the three groups, the immediate posttest scores were significantly higher than the delayed posttest scores, producing a small effect (OR = 2.03). The significant main effect of collocation type suggests that when collapsed across the three groups and immediate and delayed posttests, scores for the studied collocations were significantly higher than those for the unstudied collocations, producing a small effect (OR = 2.48). The significant fixed effect of the UVLT suggests that higher UVLT scores were associated with higher posttest scores, with a small effect (OR = 2.41). The fixed effect of lag to test was not statistically significant, which

suggests that the recency effect did not significantly affect learning. None of the interactions fitted into the model were statistically significant.

To examine where significant differences lay at immediate and delayed posttests, post-hoc analysis with Tukey's test was conducted (Table 7). The results showed that on the immediate posttest, for studied collocations, the collocation-spaced group significantly outperformed both the node-massed (OR = 8.41 [large effect]) and collocation-massed groups (OR = 4.85 [medium effect]). The collocation-massed group, however, failed to significantly outperform the node-massed group, producing a small effect (OR = 1.73). For unstudied collocations, the collocation-spaced group significantly outperformed the collocation-massed group, with a small effect size (OR = 3.13). No significant difference, however, existed between the collocation-spaced and node-massed groups (OR = 3.25), or between the collocation-massed and node-massed groups (OR = 1.04), and no more than small effects were found.

On the delayed posttest, the collocation-spaced group significantly outperformed other groups for both studied and unstudied collocations, producing small to large effect sizes (3.19 ≤ OR ≤ 6.96). The difference between the two massed groups, however, was not statistically significant for either studied (OR = 2.16 [small effect]) or unstudied collocations (OR = 1.23 [negligible effect]). The findings suggest the following order on the verb-filling test:

Studied collocations
Immediate and delayed posttests: collocation spaced > collocation massed = node massed
Unstudied collocations
Immediate posttest: collocation spaced ≥ node massed; collocation spaced > collocation massed; node massed = collocation massed
Delayed posttest: collocation spaced > collocation massed = node massed

Discussion

The present study was the first attempt to examine the effects of spacing on the knowledge of both studied and unstudied L2 collocations. Hypothesis 1 predicted an advantage of the collocation-spaced schedule over the two massed schedules for the retention of studied collocations. It was shown that the collocation-spaced schedule led to better retention of studied collocations than the massed schedules, regardless of type (collocation filling or verb filling) or timing of posttest (immediate or delayed), supporting Hypothesis 1. The collocation-spaced schedule led to superior retention possibly because it was the only condition that involved spaced retrieval practice of individual collocations. In other words, whereas retrieval opportunities for a given collocation were concentrated into a single session in the two massed schedules, they were distributed over 3 weeks in the collocation-spaced schedule. Retrieval opportunities distributed over a long time perhaps resulted in effortful retrieval, which facilitates retention according to the desirable difficulty framework (e.g., Bjork, 1994; Suzuki et al., 2019). It should also be noted that during the treatment, the same context sentence was repeated three times, instead of using three different contexts (see the "Method" section). The repetition of the same context perhaps increased the reminding potential for studied collocations. As a result, retrieval practice in the collocation-spaced schedule was not only effortful but also successful, which facilitated retention even more (reminding theory; Benjamin & Tullis, 2010; Koval, 2022).

Table 7. Results of the post-hoc analysis for treatment: Verb-filling test

Posttest	Items	Comparisons	Estimate	95% CI	SE	z	Odds Ratio	p
Immediate	Studied	Collocation spaced vs. Node-massed	2.13	[1.44, 2.82]	0.35	6.04	8.41	<.001
		Collocation spaced vs. Collocation-massed	1.58	[0.96, 2.20]	0.32	5.00	4.85	<.001
		Collocation massed vs. Node-massed	0.55	[−0.11, 1.20]	0.34	1.63	1.73	.899
	Unstudied	Collocation spaced vs. Node-massed	1.18	[0.44, 1.93]	0.38	3.11	3.25	.079
		Collocation spaced vs. Collocation-massed	1.14	[0.49, 1.79]	0.33	3.45	3.13	.028
		Collocation massed vs. Node-massed	0.04	[−0.72, 0.80]	0.39	0.11	1.04	1.000
Delayed	Studied	Collocation spaced vs. Node-massed	1.94	[1.22, 2.66]	0.37	5.31	6.96	<.001
		Collocation spaced vs. Collocation-massed	1.16	[0.54, 1.79]	0.32	3.64	3.19	.014
		Collocation massed vs. Node-massed	0.77	[0.08, 1.47]	0.35	2.19	2.16	.560
	Unstudied	Collocation spaced vs. Node-massed	1.46	[0.64, 2.27]	0.41	3.52	4.31	.022
		Collocation spaced vs. Collocation-massed	1.24	[0.56, 1.92]	0.35	3.59	3.46	.017
		Collocation massed vs. Node-massed	0.21	[−0.63, 1.06]	0.43	0.50	1.23	1.000

A limited advantage of the collocation-massed schedule over the node-massed schedule for the studied collocations was also found. On the delayed collocation-filling posttest, the collocation-massed group significantly outperformed the node-massed group, although the difference was not statistically significant on any other posttests. The limited advantage of the collocation-massed group was caused possibly by retrieval-induced facilitation (Chan et al., 2006), according to which retrieval facilitates retention of not only practiced materials but also unpracticed related materials. Specifically, in the collocation-massed group, three studied collocations for the same node verb were distributed over 3 weeks (Week 1: *draw a line*, Week 2: *draw tears*, Week 3: *draw a conclusion*). Encountering *draw a conclusion* in Week 3, for instance, might have reactivated knowledge of the two studied collocations introduced in earlier weeks (Week 1: *draw a line*, Week 2: *draw tears*), resulting in retrieval-induced facilitation from later weeks to earlier weeks. In the node-massed group, in contrast, three studied collocations for the same node verb were concentrated into a single day. As a result, retrieval-induced facilitation across weeks was not possible. At the same time, the advantage of the collocation-spaced group over the collocation-massed group suggests that effects of retrieval-induced facilitation were rather limited and repeating the same collocations across multiple sessions facilitates retention more than repeating different collocations with the same node.

The results of this study regarding Hypothesis 1 (i.e., collocation spaced > node massed = collocation massed) are consistent with those of Macis et al. (2021, Experiment 2), which showed that for intentional learning, studying collocations over multiple days facilitated learning, relative to massing them into a single day. However, this study's results were not consistent with Snoder (2017), who found that long spacing did not facilitate the retention of studied collocations. The inconsistent findings may be due to the amount of spacing used in the studies. Specifically, whereas Macis et al. (2021) compared spacing and massing (no spacing) and examined the spacing effect as in the present study, Snoder (2017) compared effects of two spacing schedulers (i.e., relatively short vs. relatively long intervals) and examined the lag effect.

Hypothesis 2 predicted that for knowledge of unstudied collocations, the collocation-spaced schedule will be the most effective, and the collocation-massed schedule will be the least. Although results on the verb-filling test showed the advantage of the collocation-spaced schedule over the other two, no significant difference was found between the two massed schedules (collocation spaced > node massed = collocation massed). Hypothesis 2, therefore, was only partially supported. The findings suggest that the benefits of spacing apply not only to recall of previously presented materials (i.e., studied words) but also to induction. The collocation-spaced schedule was the most effective for unstudied collocations possibly because participants encountered multiple collocations for the same node word every week throughout the treatment. For instance, in the first treatment session (Wednesday in Week 1; see Figure 1), participants in the collocation-spaced group were exposed to three collocations for the node verb *carry* (e.g., *carry a product*, *carry a tune*, *carry weight*). This may have allowed participants to make generalizations about what kinds of nouns the node verb could take as an object, allowing them to transfer the knowledge of studied collocations to unstudied collocations. Furthermore, the collocation-spaced group encountered the same three collocations in the subsequent 2 weeks (Wednesday in Weeks 2 and 3). The retrieval opportunities for the multiple collocations for the same node verb distributed over the 3 weeks perhaps consolidated the learners' understanding of the meaning potential of *carry*, resulting in the largest gains in the collocation-spaced group for the unstudied collocations.

In the node-massed schedule, participants were also exposed to multiple collocations for the same node word within the same day, as in the collocation-spaced schedule. This may have allowed learners to reconceptualize their knowledge of the meaning potential of the node word. However, unlike the collocation-spaced schedule, in the node-massed schedule, encounters with a given node verb were concentrated into a single session, and they were never repeated in subsequent sessions. As a result, in the node-massed schedule, learners' knowledge of the meaning potential of the node words perhaps decayed by the time of the posttests, resulting in the lack of significant difference between the two massed schedules. These findings highlight the value of distributed retrieval practice not only for studied but also for unstudied collocations.

At the same time, this study did not use a comparison group where multiple collocations for the same node word were repeated on different days over multiple weeks (e.g., *carry a product* is repeated on Mondays, *carry a tune* is repeated on Wednesdays, and *carry weight* is repeated on Fridays over 3 weeks, instead of all three collocations for *carry* being repeated on Wednesdays). As a result, it is not clear to what extent the superiority of the collocation-spaced schedule was due to the fact that participants encountered multiple collocations for the same node word on the same day. In future research, it would be useful to include a condition where multiple collocations for the same node are repeated on different days over multiple weeks.

Results of this study suggested that the collocation-spaced schedule was more effective than the two massed schedules not only for studied but also unstudied collocations. The collocation-spaced group, at the same time, may have resulted in more over-extension errors (i.e., erroneously using a target node verb to collocations where a different verb should have been used) than the other two groups. To examine whether this was the case, an error analysis was conducted. The error analysis indicated that the collocation-spaced schedule resulted in more over-extension errors than the two massed schedules (for details, see Appendix S8 in the Online Supplementary Materials). The findings suggest that although the collocation-spaced schedule enabled learners to transfer the knowledge of studied collocations to novel, unstudied collocations, it can be a double-edged sword in the sense that it may lead to over-extension errors. In future research, it may be useful to examine how over-extension errors may be reduced.

In this study, all three groups showed improvements on the verb-filling test for unstudied items on the posttests. The findings suggest that learners were able to transfer the knowledge of studied collocations to unstudied collocations. One explanation for the findings is that exposure to multiple collocations with the same node allowed learners to make comparisons between their existing knowledge of the verb's semantics and range of different uses of the verb in given collocations, which triggered semantic restructuring (Jiang, 2004). Another explanation is that learners produced novel collocations based on L1 translations of studied collocations. Some studied and unstudied collocations for a given node shared the same L1 translation. For instance, *cut* in both *cut class* (studied collocation) and *cut school* (unstudied collocation) is translated into the same Japanese word, *saboru*. Similarly, the node verbs for the following studied and unstudied collocations share the same L1 translations: *draw tears* and *draw laughs* (*sasou*), *draw attention* and *draw a line* (*hiku*), and *run an article* and *run a story* (*keisaisuru*). For these collocations, learners might have been able to guess the correct node word based solely on the L1 translations, without understanding the core meaning of the node.

To examine the effects of overlap of L1 translations among studied and unstudied items, a follow-up analysis was conducted. The follow-up analysis that included overlap

of L1 translations as fixed and random effects suggested that unstudied items that shared L1 translations with studied items were more likely to be answered successfully than those that did not share L1 translations ($p = .017$, OR = 1.51 [negligible effect]; full results of the follow-up analysis are presented in Appendix S9 in the Online Supplementary Materials). At the same time, it should be noted that inferences based on L1 translations were probably not always successful because, for some target collocations, different node verbs shared the same L1 translations. For instance, in all the following collocations, the node verbs are translated into the same Japanese word, *suru*: *cut a deal*, *make a mention*, *meet death*, *put emphasis*, and *take pains*. Because all these collocations required different node verbs, it would not be possible to guess the correct node for these collocations based solely on L1 translations (*suru*), and at least some understanding of the meaning potential of the node might have been necessary.

Although all three groups showed improvements on the verb-filling posttest for unstudied items, the posttest scores for the unstudied collocations were much lower than those for the studied collocations in all three groups (Table 3). The relatively low scores for the unstudied items may be in part due to three factors. First, during the treatment, learners were exposed to only three collocations per node. This may have made it difficult for learners to notice the core meaning underlying different uses of the node word. Second, in this study, target collocations were determined so that the choice of the node verb can be explained by the core meaning of the node verb (see Appendix S4 in the Online Supplementary Materials). At the same time, for some collocations (e.g., *meet a need*, *run an article*, *take root*), the relationship between the meaning of the collocation and the core meaning of the node verb might have been difficult to understand. This was perhaps another factor responsible for the relatively low scores for the unstudied collocations on the verb-filling test. Third, some node verbs had similar core meanings. For instance, as shown in Appendix S4, the core meanings of four node verbs (*draw*, *run*, *meet*, and *take*) involved moving something. Due to the similarity among these node verbs, learners might have had difficulty in transferring the knowledge of studied collocations to unstudied collocations, resulting in the relatively low scores for the unstudied collocations on the posttests.

Results for the quiz given at the end of each treatment session (Stage 7) have shown that the node-massed (96.4%) and collocation-massed groups (94.4%) outperformed the collocation-spaced group (75.8%) during the learning phase. The results may be partly attributed to the number of exposures before the quizzes. Specifically, whereas a quiz was given after six exposures to each collocation in the two massed groups in all 3 weeks, it was given after two (Week 1), four (Week 2), or six exposures (Week 3) in the collocation-spaced group (see Appendix S5 in the Online Supplementary Materials). On the posttests, however, the collocation-spaced group significantly outperformed the massed groups. The findings are consistent with the desirable difficulty framework (Bjork, 1994; Suzuki et al., 2019), according to which a condition that increases learning phase performance does not necessarily lead to better long-term retention than a condition that decreases learning phase performance.

This study also showed wide gaps between the learning phase and posttest performance for the two massed groups. Although the average score for the node-massed group was 96.4% on the quiz given at the end of the learning phase (Stage 7), it dropped to 14.9% and 6.6% for the immediate and delayed collocation-filling posttests, respectively. Similarly, the collocation-massed group showed a substantial decrease from the learning phase performance (94.4%) to the posttest performance (immediate collocation filling: 23.0%; delayed collocation filling: 15.9%). Figures 2 and 3 also indicate that some participants in the two massed groups scored 0 on the posttest. The significant

decrease was caused possibly because the two massed groups encountered only three collocations per treatment session, whereas the collocation-spaced group encountered nine collocations each day (Figure 1). The larger number of collocations practiced each day perhaps increased retrieval effort required for the collocation-spaced group. In other words, although retrieval practice for the two massed groups was highly successful, it was perhaps not very effortful. This may be partly responsible for the substantial decrease from the learning phase to the posttest in these two groups.

Although direct comparisons of this study and other studies are difficult due to a number of methodological differences, posttest scores in this study were relatively high, compared with other studies involving L2 collocation learning. Boers, Demecheleer, et al. (2014), for instance, report 4.5% to 11.2% gains on the verb-filling posttest and 8.9% to 13.7% gains on the collocation-filling posttest, after a single treatment session. These scores were much lower than those obtained by the collocation-spaced group in this study (62.1% on the immediate verb-filling and 57.8% on the immediate collocation-filling posttest). The results may demonstrate the value of spacing for collocation learning. As a case in point, Ferguson et al. (2021) found that a treatment that involved the repetitions of same collocations three times after 2-day gaps led to gains that are similar to or larger than those in this study (48.0% to 62.5% on the immediate and 38.0 % to 64.7% on the delayed posttests).

Pedagogical implications

The findings of this study suggest that introducing spacing in terms of individual collocations (i.e., collocation-spaced schedule) facilitates the knowledge of both studied and unstudied collocations. Pedagogically, the findings suggest that it may be useful for learners to be exposed to multiple collocations containing the same node regularly. This study also showed that although the two massed groups significantly outperformed the collocation-spaced group during the learning phase, the collocation-spaced group resulted in higher posttest scores than the massed groups. Pedagogically, the findings suggest that learners or instructors should not be discouraged even if the treatment induces a large number of incorrect responses during learning (desirable difficulty framework).

Concluding remarks

Although many studies have examined the effects of spacing on vocabulary learning, most studies have investigated the learning of single words. Related studies that compared massing and spacing for collocation learning so far investigated the learning of only one collocate per node word (Macis et al., 2021; Snoder, 2017). Thus, it was not clear how the spacing of multiple collocations with the same node affects the knowledge of studied and unstudied collocations. The findings of this study are valuable because they suggest that introducing spacing in terms of individual collocations (collocation-spaced schedule) may facilitate the knowledge of both studied and unstudied L2 collocations. At the same time, because this study was the first to examine the role of spacing for collocation learning in this way, it has several limitations.

First, this study was conducted within an authentic classroom setting. Although classroom-based research helps increase ecological validity and has its benefits (Rogers & Cheung, 2021), it is also limited in that experimental manipulations are not as tightly controlled as in laboratory studies. For instance, during the treatment in this study,

participants were asked to say the correct answers aloud (Stages 3–6 during the treatment; see Appendix S5 in the Online Supplementary Materials). Although over-hearing other students' responses is common in real-world classroom settings, it might have affected learning. In future research, it may be useful to replicate this study in laboratory settings. Second, in this study, the collocation-spaced group, who showed better learning outcomes than the two massed groups, had the highest UVLT scores among the three groups. Although the UVLT score was used as a covariate in the analysis to control for English proficiency effects, in future research, it may be useful to compare groups that are equivalent in their proficiency levels. Considering the value of collocational knowledge for the appropriate and fluent use of L2 vocabulary, further research examining the effects of spacing on collocation learning is warranted. Investigating the effects of spacing on the knowledge of unstudied collocations is also valuable from a theoretical viewpoint because it allows researchers to examine whether the benefits of spacing apply not only to recall of previously presented materials but also to induction.

References

Ackermann, K., & Chen, Y. (2013). Developing the academic collocation list (ACL): A corpus driven and expert-judged approach. *Journal of English for Academic Purposes*, *12*, 235–247.

Bahrick, H. P., & Phelps, E. (1987). Retention of Spanish vocabulary over 8 years. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *13*, 344–349.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48.

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

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). MIT Press.

Boers, F., Dang, T. C., & Strong, B. (2017). Comparing the effectiveness of phrase-focused exercises. A partial replication of Boers, Demecheleer, Coxhead and Webb (2014). *Language Teaching Research*, *21*, 362–380.

Boers, F., Demecheleer, M., Coxhead, A., & Webb, S. (2014). Gauging the effects of exercises on verb–noun collocations. *Language Teaching Research*, *18*, 54–74.

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

Carpenter, S. K., & Mueller, F. E. (2013). The effects of interleaving versus blocking on foreign language pronunciation learning. *Memory and Cognition*, *41*, 671–682.

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

Chan, J. C. K., McDermott, K. B., & Roediger, H. L. (2006). Retrieval-induced facilitation: Initially nontested material can benefit from prior testing of related material. *Journal of Experimental Psychology: General*, *135*, 553–571.

Chen, H., Cohen, P., & Chen, S. (2010). How big is a big odds ratio? Interpreting the magnitudes of odds ratios in epidemiological studies. *Communications in Statistics: Simulation and Computation*, *39*, 860–864.

Conklin, K., & Carrol, G. (2018). First language influence on the processing of formulaic language in a second language. In A. Siyanova-Chanturia & A. Pellicer-Sánchez (Eds.), *Understanding formulaic language. A second language acquisition perspective* (pp. 62–77). Routledge.

Durrant, P., & Schmitt, N. (2010). Adult learners' retention of collocations from exposure. *Second Language Research*, *26*, 163–188.

Eyckmans, J., Boers, F., & Lindstromberg, S. (2016). The impact of imposing processing strategies on L2 learners' deliberate study of lexical phrases. *System*, *56*, 127–139.

Ferguson, P., Siyanova-Chanturia, A., & Leeming, P. (2021). Impact of exercise format and repetition on learning verb–noun collocations. *Language Teaching Research*. Advance online publication. https://doi.org/10.1177/13621688211038091

González-Fernández, B., & Schmitt, N. (2015). How much collocation knowledge do L2 learners have? The effects of frequency and amount of exposure. *International Journal of Applied Linguistics*, *166*, 94–126.

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

Henriksen, B. (2013). Research on L2 learners' collocational competence and development: A progress report. In C. Bardel, C. Lindqvist, & B. Laufer (Eds.), *L2 vocabulary acquisition, knowledge and use: New perspectives on assessment and corpus analysis* (pp. 29–56). EuroSLA.

Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge University Press.

Jiang, N. (2004). Semantic transfer and its implications for vocabulary teaching in a second language. *The Modern Language Journal*, *88*, 416–432.

Kang, S. H. (2016). The benefits of interleaved practice for learning. In J. C. Horvath, J. M. Lodge, & J. Hattie (Eds.), *From the laboratory to the classroom: Translating science of learning for teachers* (pp. 79–93). Routledge.

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

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the "enemy of induction"? *Psychological Science*, *19*, 585–592.

Koval, N. G. (2022). Testing the reminding account of the lag effect in L2 vocabulary learning. *Applied Psycholinguistics*, *43*, 1–40.

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

Lenth, R. (2021). *Emmeans: Estimated marginal means, aka least-squares means. R package* (Version 1.7.1-1) [Computer software]. https://CRAN.Rproject.org/package=em means

Macis, M., Sonbul, S., & Alharbi, R. (2021). The effect of spacing on incidental and deliberate learning of L2 collocations. *System*, *103*, 102649. https://doi.org/10.1016/j.system.2021.102649

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*, 287–311.

Nakata, T., & Webb, S. (2016). Does studying vocabulary in smaller sets increase learning? The effects of part and whole learning on second language vocabulary acquisition. *Studies in Second Language Acquisition*, *38*, 523–552.

Paquot, M., & Granger, S. (2012). Formulaic language in learner corpora. *Annual Review of Applied Linguistics*, *32*, 130–149.

Peters, E. (2016). The lexical burden of collocations: The role of interlexical and intralexical factors. *Language Teaching Research*, *20*, 113–138.

Pellicer-Sánchez, A. (2017). Learning L2 collocations incidentally from reading. *Language Teaching Research*, *21*, 381–402.

Plonsky, L., & Derrick, D. J. (2016). A meta-analysis of reliability coefficients in second language research. *The Modern Language Journal*, *100*, 538–553.

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

R Core Team. (2021). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

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, 1138–1156.

Shin, D. (2006). A collocation inventory for beginners. *Unpublished doctoral dissertation*. Victoria University of Wellington, New Zealand.

Siyanova, A., & Schmitt, N. (2008). L2 learner production and processing of collocation: A multi-study perspective. *The Canadian Modern Language Review*, 64, 429–458.

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

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

Suzuki, Y., Nakata, T., & DeKeyser, R. M. (2019). The desirable difficulty framework as a theoretical foundation for optimizing and researching second language practice. *The Modern Language Journal*, 103, 713–720.

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.

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, 245–265.

Toomer, M., & Elgort, I. (2019). The development of implicit and explicit knowledge of collocations: A conceptual replication and extension of Sonbul and Schmitt (2013). *Language Learning*, 69, 405–439.

Webb, S., & Chang, A. C.-S. (2022). How does mode of input affect the incidental learning of collocations? *Studies in Second Language Acquisition*, 43, 55–77.

Webb, S., & Kagimoto, E. (2009). The effects of vocabulary learning on collocation and meaning. *TESOL Quarterly*, 43, 55–77.

Webb, S., Newton, J., & Chang, A. C.-S. (2013). Incidental learning of collocation. *Language Learning*, 63, 91–120.

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

Wood, D. (2020). Categorizing and identifying formulaic language. In S. Webb (Ed.), *Routledge handbook of vocabulary studies* (pp. 30–45). Routledge.