Improving English Learners' Productive Collocation Knowledge: The Effects of Involvement Load, Spacing, and Intentionality

Per Snoder

This article reports on a classroom-based experiment that tested the effects of three vocabulary teaching constructs (involvement load, spacing, and intentionality) on the learning of English verb-noun collocations—for example, "shelve a plan." Laufer and Hulstijn's (2001) "involvement load" predicts that the higher the motivational-cognitive load of a task, the more effectively it promotes word retention. "Spacing" refers to the advantage of spreading out learning opportunities for words as opposed to massing them. "Intentionality" comprises two word processing modes: intentional learning (posttest announced) and incidental learning (posttest unannounced), where the former is claimed to outperform the latter. The constructs were integrated into an intervention study with 59 adolescent L1 Swedish learners of English in within- and between-subjects designs. Learners processed target items three times when performing tasks that operationalized the constructs. Three posttests of productive knowledge of target items were administered. Statistical analyses of gain scores show that neither involvement load nor spacing had a significant positive impact on learning gains. Significant effects were found on three measures for intentional learning when compared to incidental learning. The findings are discussed in relation to previous research and their implications for English language teaching (ELT).

Cet article rend compte d'une expérience basée en salle de classe et qui évalue les effets de trois concepts pédagogiques en enseignement du vocabulaire (charge d'implication, espacement et intentionnalité) sur l'apprentissage d'expressions figées en anglais consistant d'un verbe et d'un nom. La « charge d'implication » de Laufer et Hulstijn (2001) prédit que plus la charge motivationnelle et cognitive d'une tâche est élevée, plus elle favorise la rétention des mots. « L'espacement » fait référence à l'avantage de répartir les occasions d'apprentissage des mots plutôt que de les rassembler. « L'intentionnalité » comprend deux modes de traitement des mots: l'apprentissage intentionnel (post-test annoncé) et l'apprentissage accidentel (post-test pas annoncé), le premier supposément donnant de meilleurs résultats que le second. Ces concepts ont été intégrés dans une étude d'intervention impliquant 59 adolescents suédois qui apprenaient l'anglais et suivant un modèle à mesures répétées et un modèle inter-sujets. Les participants ont traité des éléments cibles trois fois en accomplissant des tâches qui opérationnalisaient les concepts. Trois post-test portant sur des connaissances productives des éléments cibles ont eu lieu. Des analyses statistiques des notes ont indiqué que ni la charge d'implication ni l'espacement n'avaient eu un impact positif significatif sur les notes. Trois me-

VOLUME 34, ISSUE 3, 2017 PP. 140–164
http://dx.doi.org/10.18806/tesl.v34i3.1277

sures de l'apprentissage intentionnel comparé à l'apprentissage accidentel ont révélé des effets significatifs. Nous discutons des résultats par rapport à la recherche antérieure et en fonction de leurs incidences sur l'enseignement de l'anglais.

KEYWORDS: English language teaching (ELT), collocation learning, involvement load, spacing, intentionality

This study examines how second language (L2) instruction can facilitate the learning of formulaic sequences (FSs) in classroom settings. It is widely acknowledged that native speakers are able to operate fluently in the language with the help of a large repertoire of FSs that serve cognitive and social functions (Pawley & Syder, 1983; Wray, 2002). In English, these recurring (semi-) fixed word combinations include idioms (*steal the show*), phrasal verbs (*drop out*), and collocations (*run a shop*), to mention a few well-known categories. FSs are also important for L2 learners. However, descriptive research has shown that learners at all proficiency levels struggle with the appropriate use of FSs. There is no shortage of intervention studies that have tested different ways in which L2 instruction may increase learners' formulaic competence. Even so, useful guidelines for instructors are scarce (for reviews see Boers & Lindstromberg, 2012; Meunier, 2012). The present study attempts to fill this knowledge gap by investigating what English language teaching (ELT) practitioners can do to improve learners' productive knowledge of collocations. The investigation focused on the effects of three vocabulary teaching constructs: Laufer and Hulstijn's (2001) *involvement load*, *spacing*, and *intentionality*, relating in turn to task quality, intervals between exposures to target items, and whether posttests are announced or not. The participants in the study were L1 Swedish adolescent learners of L2 English.

Literature Review

English Verb-Noun Collocations

The study adopts Henriksen's (2013) definition of collocation, which combines the frequency-based and phraseological views of collocation:

> Collocations are frequently recurring two-to-three word syntagmatic units which can include both lexical and grammatical words, e.g., verb + noun (*pay tribute*), adjective + noun (*hot spice*), preposition + noun (*on guard*), and adjective + preposition (*immune to*). (p. 30)

This definition is relevant for classroom L2 research as it includes collocations that are both common and meaningful (see Sonbul, 2012, pp. 56–63, for discussion). The study targets English verb-noun collocations such as *carry a risk*, which play an important role for two reasons: they are the most frequent type of collocation and they convey the gist of messages (Howarth, 1998).

Verb-noun collocations are also problematic for learners, who have difficulties in choosing which English verb to combine with the noun in productive use. Many errors are due to interlingual incongruence, a lack of L1-L2 translational overlap that may lead to non-native forms such as **make homework* (Nesselhauf, 2005) and **solve the disease* (Laufer and Waldman, 2011). A recent study by Peters (2016) produced two relevant results: incongruent target collocations were harder to learn than congruent ones, and verb-noun collocations were harder to learn than adjective-noun collocations.

The challenges English collocations present for learners have generated a number of intervention studies and suggested teaching methods. Lewis's oft-cited textbook *Teaching Collocation* (2000) advocates a shift in ELT practices toward raising learner awareness of collocations in input encountered outside school. Although qualitative data have indicated that awareness-raising techniques may promote collocation learning (Ying & O'Neill, 2009), a review of intervention studies showed that they are unlikely to lead to substantial learning gains (Boers & Lindstromberg, 2012, p. 99). The latter authors also question Lewis's (2000) overreliance on learners' capacity of noticing and learning collocations independently, based on three factors: (a) identifying collocations requires expert help from teachers or learning resources, (b) it overestimates learners' willingness to engage in such activities extramurally, and (c) it disregards the role of memory in internalizing unknown language "chunks" (Boers & Lindstromberg, 2009, pp. 19–21).

This line of argument implies that instructors should be responsible for improving learners' collocational knowledge, but useful guidelines are still scarce. Particular interest has been given to enhanced frequency of target items in "flooded" instructional materials. However, this contrived frequency poses a threat to the ecological validity of the study: how often do instructional materials contain 15 occurrences of the same verb-noun collocations as in Webb, Newton, and Chang's (2013) study of graded readers? Other studies have investigated collocation learning via computer-generated concordances—where target items are highlighted in sentential contexts—but their effectiveness remains uncertain. While Chan and Liou (2005) report sizeable short-term gains from using a concordancer compared to a more deductive teaching method, the advantage was lost in delayed posttesting. Wu, Witten, and Franken (2010) focused on the use of a concordancer to assist learners in correcting malformed collocations, which led to promising results as learners correctly replaced 67% of them. Posttests were not administered, and evidence of learning is therefore lacking. Liu (2010) suggests using a cognitive analysis of collocation in concordances and speculates that it offers opportunity for processing and raises learners' awareness of collocation (pp. 26–27). Even if empirical research finds support for Liu's approach, it may not be relevant due to generic problems of computer-assisted language learning (CALL), such as lack of proper equipment and adequate training for instructors (McEnery & Xiao, 2011).

Form-focused instructional approaches to collocation learning have demonstrated partly promising results. In two studies by Laufer and Girsai (2008) and Szudarski (2012), learners who were explicitly taught the target items outperformed their peers in a more meaning-focused learning condition. Form-focused instruction was operationalized by contrastive analysis and translation in the former study and various written exercises drawing learners' attention to target item form in the latter. However, a study on the effects of written exercises on collocation learning produced the opposite result (Boers, Demecheleer, Coxhead, & Webb, 2014). Not only were the written exercises ineffective in boosting learning gains, the "matching exercise" format led to sizeable unlearning. A case in point is the learner who correctly produced *take approach* at the pretest, but incorrectly replaced the verb component with **give* at the posttest. The authors attribute this erroneous cross-association to the matching exercise format in the treatment, where the three options *give*/*run*/*take* were presented to learners. Their recommendation is that collocations be presented as intact wholes (Boers et al., 2014, p. 68).

What, then, can instructors do to facilitate learning of English collocations when awareness-raising, flooded instructional materials, and CALL have proven either ineffective or problematic? It seems that setting learners written form-focused tasks where incongruent verb-noun collocations are processed as intact wholes may be beneficial, but more empirical research is needed to find the most effective procedures. The following sections review the literature on three vocabulary teaching constructs that are deemed relevant for empirical investigation to address this issue: Laufer and Hulstijn's (2001) involvement load, spacing, and intentionality. These constructs were selected for the study as they lend themselves easily to manipulation by the instructor, are adaptable to suit learners of most proficiency levels (thus relevant for both second and foreign language learning settings), and have so far only been considered in the context of single words.

Involvement Load

Laufer and Hulstijn's (2001) Involvement Load Hypothesis (ILH) makes specific predictions about the relative effectiveness of different classroom tasks on incidental (nonintentional) L2 vocabulary learning. The central idea of the ILH is that learning new words is a function of the motivational and cognitive "involvement load" induced by tasks: the higher the involvement load of a task, the more effectively it will promote learning gains. Three components make up the involvement load: *need*, *search*, and *evaluation*. Each component is given a score when operationalized to calculate an involvement load index of maximum 5 points: 0–2 for need and evaluation, and 0–1 for search.

Several intervention studies have tested the ILH and compared three tasks inducing different involvement loads. Most studies found support for it in

the sense that tasks with a higher involvement load generated greater learning gains than tasks with a lower one (e.g., Keating, 2008; Kim, 2008). With the exception of Hulstijn and Laufer (2001), ILH studies have only investigated the learning of single words.[1] Some studies failed to provide full support for the ILH, in that tasks with different involvement loads yielded the same learning gains (Hulstijn & Laufer, 2001; Nassaji & Hu, 2012; Zou, 2017). Authors of studies that found support for the ILH still question its validity, claiming that it does not control for instructional materials nor learner behaviour during task performance. Such confounding factors include whether or not learners actually looked at the glossed target items (Eckerth & Tavakoli, 2012). Lee and Pulido (2017) focused on the motivation (i.e., "need") component of involvement load for incidental L2 vocabulary gain from reading two texts. They found that, all other textual aspects being equal, the text that their learners rated as highly interesting was significantly more effective in fostering word learning than the text rated as less interesting.

All but one of these ILH studies used a between-subjects design in which each participant performed one task only and mean posttest scores were compared. This implies that confounding learner variables such as language proficiency were not controlled for. In one study by Eckerth and Tavakoli (2012), participants performed all three tasks to control for differences between treatment materials (texts), not individual learner characteristics. However, the tasks and immediate posttests of that study were spread out over a 3-week period during which participants are likely to have expected posttests of target items and "switched on" an intentional learning mode. The two authors address this issue (2012, p. 245) by questioning the incidental-intentional learning dichotomy, arguing that vocabulary learning is always intentional from the learner's perspective, and that a continuum model is more adequate.

Previous ILH studies have targeted single words and not FSs (not considering Hulstijn & Laufer, 2001), but there are several reasons for testing the ILH on FSs. First, the learning of single words and that of FSs have been argued to be essentially similar, both occurring incrementally (Schmitt & Carter, 2004, p. 6). Second, Hulstijn and Laufer (2001, p. 553) suggest that tasks with higher involvement loads may be required to learn challenging words, and they mention a category of FSs, namely idioms. Third, L2 vocabulary researchers have called for more studies on effective ways of teaching FSs, particularly collocations (e.g., Nesselhauf, 2005, p. 3; Webb & Kagimoto, 2011, pp. 259–260). Importantly, future studies testing the ILH on FSs should take the methodological issues raised above into consideration to strengthen the validity of findings. The solutions to these issues include the use of a within-subjects design, ascertaining that participants are unaware of upcoming posttests, and producing teaching materials that are highly motivating for the target age group.

Spacing

Spacing effects in learning and memory research refer to the demonstrated robust retention advantages of spreading out repetitions of items of learning instead of focusing the repetitions on one single intensive study period, so-called *massed* learning. Spacing effects have been found in several hundred empirical studies in cognitive and educational psychology in various contexts and with different types of learning material, including vocabulary (see Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006, for review).

Studies of spacing effects in L2 vocabulary learning have targeted single words in so-called paired-associate learning (PAL) tasks where researchers manipulate presentations and cued retrievals of target items between learners' L1 and the L2 in question. Nakata (2015) and Schuetze (2015) are recent PAL studies that demonstrated the advantages of *expanding* spaced learning schedules, where intervals between exposures to target items are gradually increasing, over *equally/uniformed* spaced learning schedules where intervals are kept constant.

One aspect of spacing effects in L2 vocabulary learning that has not been investigated is whether a spaced learning schedule, whether expanding or equal, is more effective than an *intensive* learning schedule. Intensive learning refers to a schedule in which exposures to target items recur repeatedly on a daily basis rather than occurring during one single study period as in massed learning. The pedagogical relevance of such intensive learning is that it emulates real-life classroom conditions in which learners typically have two to three lessons per week, often—by necessity—on consecutive days. In addition, *pure* massed learning, occurring during one single occasion, has been claimed to be unrealistic (Seabrook, Brown, & Solity, 2005).

Another feature in most spacing studies on L2 vocabulary learning is the fact that they are laboratory-based (e.g., Nakata, 2011, 2015), which raises questions of generalizability for mainstream L2 pedagogy. Though some studies are set in the classroom (e.g., Schuetze, 2015), they are based entirely on computer-delivered treatment and not regular teacher-led instruction. It is therefore relevant to conduct a spacing study set in an authentic classroom setting where the treatment consists of teacher-led instruction and that is thus applicable to educational settings (Kornell, 2009). Furthermore, no L2 vocabulary learning study so far has examined spacing effects for lexical items larger than single words and such items include FSs. Given that simple and complex lexical items are assumed to be learned incrementally (Schmitt & Carter, 2004, p. 6), it remains to be investigated whether spacing effects occur for both types.

Intentionality

The degree of intentionality in the learning of new material is a complex issue, as it involves nonobservable cognitive phenomena (for discussions see

Hulstijn, 2001, 2003). A useful distinction for L2 classroom research drawn by Doczi and Kormos (2016, p. 120) is that between learning *conditions* and learning *processes*. The learning conditions are controlled by the instructor who may consciously induce various degrees of intentional learning on learners. The learning processes, by contrast, are controlled by the individual learner and are likely to fluctuate on an incidental-intentional continuum, depending on various elusive factors such as attention.

The present study draws on this distinction and operationalizes intentionality in Eysenck's (1982, p. 198) terms as "prelearning instructions that either do, or do not, forewarn subjects about the existence of a subsequent retention test." Importantly, the term "intentional learning" is thus only used to denote this specific procedure and not the general meaning of the term as in "explicit" or "form-focused instruction" used in other collocation learning studies. Recently, intervention studies of collocation learning have used incidental learning conditions designs with unannounced posttests and found mixed effects of various variables, such as input flood (Pellicer-Sánchez, 2017; Peters, 2014; Szudarski & Carter, 2016; Webb et al., 2013). To date, no study has investigated the effects of intentional learning conditions of collocations compared to incidental ones. This is surprising, considering the beneficial effects intentional, sometimes termed "deliberate," learning has demonstrated for single L2 words (e.g., Elgort, 2011). There are many potential advantages of isolating target collocations in a glossed (L1-L2) list and asking learners to memorize them for an immediate test: (a) it disambiguates their meaning, (b) it makes them more noticeable, (c) it gives learners the choice of strategy to use for memorizing them, and (d) it is practical for instructors. This study attempts to investigate whether and to what extent these assumptions are borne out empirically.

Research Questions

The present study set out to examine what instructors can do to facilitate learning of English collocations in classroom settings. The three vocabulary teaching constructs—involvement load, spacing, and intentionality—were integrated into the research design. The following three research questions (RQs) were formulated to guide the investigation:

RQ1: Do collocation tasks with a higher involvement load consistently generate higher learning gains of target collocation than tasks with a lower involvement load?

RQ2: Is an expanding spaced learning schedule more effective in facilitating learning gains of target collocations than an intensive learning schedule?

RQ3: What are the effects of intentional learning during the third exposure to target collocations compared to a third incidental encounter?

Method

Design of Study

A quasi-experimental study was conducted in a Swedish high school to collect data from two intact classes of learners, hereafter "Class 1" (n = 30) and "Class 2" (n = 29). The study used a pretest/treatment/posttest design. During the three treatments 1–3, the participants performed seven written tasks (A–G) that focused on the target items, for example fill-in-the gap and writing new titles for texts. These tasks operationalized the three vocabulary teaching constructs under examination. The effects of the treatment were measured through three posttests tapping participants' productive knowledge of target items. Table 1 presents an overview of the study.

Table 1
Study Overview

Sessions	Classroom activities	Posttest	Construct(s) in focus
Treatment 1	Participants performed Tasks A–D focusing on the target collocations	Yes: unannounced	- Involvement load: Tasks A–D differed in this respect - Spacing, 1st exposure
Treatment 2	Participants read six short texts containing the target collocations and answered questions on the texts: Task E	No	Spacing, 2nd exposure
Treatment 3	- First, participants reread three short texts containing 14 of the target collocations and wrote new titles for these texts: Task F - Second, participants studied the 14 other target collocations for an announced posttest: Task G	Yes: unannounced for 14 of the target collocations in the first part of the treatment and announced for the other 14 in the second part	- Spacing, 3rd exposure - Intentionality: unannounced and announced posttests
Delayed posttest (3 weeks later)	—	Yes: unannounced	—

Participants

The participants were L1 Swedish learners of English aged 16. Fifty-nine participants attended Treatment 1 where the ILH was tested and 14 participants dropped out over the following sessions due to absence from school, a recurring feature of classroom research (Schmitt, 2010, p. 150). All participants

had received a passing grade or higher in English when graduating from compulsory school prior to the study, which corresponds to Level B1.1 "independent user" in the CEFR (Council of Europe, 2001). In a pre-experimental questionnaire, most participants reported having extensive contact with English outside school. Such extramural exposure is held to explain why many Swedish learners of English have quite advanced levels of receptive skills in English (Gyllstad, 2007). Considering this circumstance, it was less relevant to tap into receptive knowledge of target items in the study.

Target Items

The target items were 28 collocations, hereafter referred to as the "target collocations" (see Appendix A). Their inclusion was based on six criteria:

1. Phraseological category: verb-noun collocations
2. L1-L2 incongruence
3. Pretest unfamiliarity below 10% (see section on data collection)
4. Occurrence in the *Oxford Collocations Dictionary for Students of English* (OCDE; McIntosh, Francis, & Poole, 2009)
5. Occurrence in a bilingual (L1-L2/L2-L1) dictionary (Petti, 1994)
6. Occurrence in an L2 reference corpus

From a phraseological perspective it can be argued that some of the target collocations of the study are idioms rather than collocations, for example *to foot the bill*. This distinction is based on Howarth's continuum model (1998) and is a function of their compositionality. However, the inclusion criterion for collocations in the OCDE (McIntosh et al., 2009) is explicitly stated as being based on their status as collocations and not idioms (p. vi), and the target collocations are therefore treated as such. The L1-L2 incongruence of the target collocations was based on the entries in the cited bilingual dictionary with no literal translation overlaps. One example is *to stir the imagination*, translated into *egga fantasin* in Swedish in the relevant entry and not **röra om fantasin*, which is the literal translation based on the researcher's intuition as a native speaker of Swedish. The target collocations occur with various frequencies in the Corpus of Contemporary American English (COCA) as established by lemmatized searches, the only exception being *to sack employees*, which is in the British National Corpus (BNC) and marked as especially British English (McIntosh et al., 2009, p. 269).

Materials

The 28 target collocations were incorporated into the researcher-developed materials used in the treatment of the study (see Appendix B). The materials—10 nonfiction texts and written tasks—were produced in collaboration with two native speakers of English. There were two reasons for creating original materials. First, occurrences of individual types of collocations are

few and far between (Webb et al., 2013, p. 95), and it was therefore not possible to conduct a study based on authentic texts. Second, it was necessary to control for exposure to the target collocations by an equal distribution in the materials. The texts were submitted to a Vocab Profile analysis using *www.lextutor.ca* to investigate whether they were of comparable linguistic difficulty in terms of how frequent the vocabulary of each text was. Two Chi-square tests were run and revealed no significant difference between the texts in this respect. The materials were piloted on and evaluated by a nonparticipant class of learners. The pilot was followed by minor modifications. The learners' evaluations indicated that the materials were appropriate for the designated age group regarding the level of linguistic difficulty of the texts and the cognitive complexity of the written tasks. The learners also found that the content of the texts had the potential to pique their interest, which has been found to positively influence vocabulary learning in ILH studies (cf. Lee & Pulido, 2017).

Procedures

Before the experiment started, participants completed a pre-experimental questionnaire.[2] Instructions to participants were given by the researcher, hereafter the "instructor," to increase the control of the experiment and to ascertain that the two classes received the same treatment. During the three treatment sessions, the instructor first gave instructions and then circulated among participants to monitor their task performance. No indications of off-task behaviour, for example failure to comply with task instructions, were observed.

a. The Involvement Load Hypothesis (ILH). The operationalizations of the ILH are presented in Table 2 below. Note how the involvement load index for each task (A–D) was implemented—for example, "*search* 0" entailed that target collocations were provided in a glossed list. Two procedures were held constant across all tasks: (a) the instructor read the target text aloud before setting the task, and (b) participants worked individually in silence. The second procedure was crucial, as learner-learner interaction risks subverting the operationalizations of the ILH. Participants performed the four tasks during one long session, and it is therefore argued that they were not aware of the unannounced posttest after the fourth task, which they may have been had the tasks been performed on separate occasions (cf. Eckerth & Tavakoli, 2012). The order of the tasks was reversed for Class 2 to control for recency effects at the immediate posttest. All treatment materials—participants' worksheets with gap-fills and original sentences—were collected after task completion to prevent undesired additional exposure and allow for analysis of task performance. A sample of 10% of participants' worksheets was scrutinized, and it was found that in 98.5% of them participants had followed the task instruction and produced acceptable answers.

Table 2
Description of Tasks A–D and Distribution of Involvement Load Index

Task	Task description	Involvement load index (need, search, and evaluation)
Task A	Participants received a gapped version of the text. They filled in the gaps by choosing from seven target collocations provided in a glossed list. Instructor displayed correct answers when participants had completed the task.	**2** (*need* 1, *search* 0, *evaluation* 1)
Task B	Participants received a glossed list of seven target collocations and wrote original sentences in which they were incorporated. An example sentence was provided.	**3a** (*need* 1, *search* 0, *evaluation* 2)
Task C	Participants received a list of seven target collocations in Swedish and searched in dictionary entries for their English translations. They filled in the gaps of a gapped version of the target text by choosing from these translations. Instructor displayed correct answers when participants had completed the task.	**3b** (*need* 1, *search* 1, *evaluation* 1)
Task D	Participants received a list of seven target collocations in Swedish and searched in dictionary entries for their English translations. They wrote original sentences in which they were incorporated. An example sentence was provided.	**4** (*need* 1, *search* 1, *evaluation* 2)

b. Spacing. Spacing was operationalized in the study through two different learning schedules. In the *expanding* learning schedule the three treatments occurred with gradually longer intervals: Day 1, Day 7, and Day 16. In the intensive learning schedule, by contrast, the intervals between the three treatments were kept as short as logistically possible: Day 1, Day 2, and Day 4. The first exposure to the 28 target collocations in Treatment 1 was the test of the ILH. The second exposure was Treatment 2, during which participants read the six short texts and answered comprehension questions. The third exposure occurred during Treatment 3 and is presented next.

c. Intentionality. Intentionality was investigated by comparing three incidental exposures to 14 of the target collocations—the "3INC" condition—with two incidental exposures followed by an intentional learning component for the other 14 target collocations—the "2INC+INT" condition. In the first half of Treatment 3, participants reread three of the texts from Treatment 2 that contained 14 target collocations and were tasked to suggest a new title for each text as it had been deleted. This procedure was their third incidental exposure to these 14 target collocations. In the second part of Treatment 3, participants were given a glossed list of the remaining 14 target collocations and were tasked to intentionally learn them for an upcoming test after 15 minutes. The instructor specified that participants should study all of them for the posttest even if they already knew some of them partially or fully.

Data Collection

The participants were not pretested on the target collocations, as it carried the risk of drawing their attention to them and the specific purpose of the study. Instead, two classes of nonparticipant learners (*N* = 44) from a high school in the same district took a 50-item pretest identical to the format in the present study. Only items known by fewer than 10% were included in the study (see Appendix A). This procedure is similar to the original involvement load study by Hulstijn and Laufer (2001), and the percentage is below that of comparable studies such as Sonbul (2012, p. 88). Three identical posttests were administered (see Table 1) that only tapped controlled productive knowledge of target collocation form, as receptive knowledge in most cases is unproblematic for the participants (see Gyllstad, 2007, with participants of the same L1 and age as in the present study). Posttests 1 and 2 were administered immediately after Treatments 1 and 3, while the delayed posttest was administered 3 weeks after Treatment 3. An example of a test item is displayed below:

Friska upp minnet __J_____ one's memory

The translation format was adapted from Laufer and Goldstein's (2004) "active recall." The participants were asked to complete the target collocation by filling in a gap in the English translation of the Swedish cue (as in the example above where the two letters "og" are missing). The first letter of the verb component was provided in order to avoid possible but undesired alternatives (cf. Peters, 2014). As argued in the literature review above, the main difficulty for learners of English lies in using the correct verb, and the noun was therefore supplied. A and B versions of the posttests were created, containing the same target collocations but listed in different orders. The two versions were alternately given to participants sitting next to each other at the posttests to minimize the risk of participants collaborating. No indication of participant collaboration was observed. The posttests were scored using the following principles. Correctly spelled answers were given one point. Incorrectly spelled answers that phonetically matched the target verb or had inflected verbs were also given one point, for example **jogg* and **jogging* for *jog*. For more complex spelling mistakes a native speaker of English was consulted. The answer was displayed in isolation, and she was asked to read it out loud. If she failed to produce the correct verb the answer received zero points, which occurred for the answer **aproace* for *approach*. Misread verbs, the wrong verb, or a blank line were given zero points. A second independent native speaker of English corrected 10% of the posttests following the scoring principles outlined above. The interrater reliability reached .97.

Results

The Involvement Load Hypothesis

Table 3 displays the descriptive statistics of the posttest scores for the test of the Involvement Load Hypothesis (ILH).

Table 3
Descriptive Statistics for the Test of the ILH

	Maximum score: 7 points		
Task	*N*	*M*	*SD*
Task A: Involvement load 2	59	3.24	2.28
Task B: Involvement load 3a	59	3.42	1.94
Task C: Involvement load 3b	59	3.14	2.19
Task D: Involvement load 4	59	2.66	2.04

Note. *N* = number of participants, *M* = mean scores, *SD* = standard deviations.

A one-way repeated measures ANOVA was run to test whether the differences between the four means were statistically significant. The results are displayed in Table 4 and include all possible comparisons between the four tasks using the LSD post-hoc test.

Table 4
Results of ANOVA with Pairwise Comparisons of Means of the Four Tasks (A–D)

Comparison of tasks		Mean difference	*p*-value
Task A	vs Task B	-.186	.345
	vs Task C	.102	.597
	vs Task D	.576*	.006
Task B	vs Task C	.288	.188
	vs Task D	.763*	.000
Task C	vs Task D	.475*	.014

*The mean difference is significant at the .05 level

As can be seen from Table 4, the only significant differences between mean scores were found between Task D and each of the three other Tasks A–C where *p*-values are below the significance level of .05. This result indicates that the task with the highest involvement load generated significantly lower learning gains than the three other tasks with relatively lower involvement loads, which runs counter to the predictions of the ILH. In addition, no significant differences were found between the means of Tasks A–C in pairwise comparisons. This result also contradicts the ILH, which predicts that Tasks B and C should be more effective than Task A.

Spacing

Table 5 displays the descriptive statistics relevant to the investigation of spacing effects in the study. The number of participants amounted to 45 due to participant dropout. A between-subject design was used in which each participant followed one learning schedule only.

Table 5
Descriptive Statistics for Spaced vs. Intensive Learning Schedules

	Maximum score: 14 points		
Learning schedule	*N*	*M*	*SD*
Spaced learning	25	7.72	3.62
Intensive learning	20	5.85	2.94

Note. *N* = number of participants, *M* = mean scores, *SD* = standard deviations.

An independent samples *t*-test was run on the mean scores for the two learning schedules. The result of the *t*-test was not statistically significant: t (43): 1.911, $p = 0.063$. This result indicates that, in this study, spacing the participants' three exposures to the target collocations was not significantly more effective in promoting learning gains than when the target collocations were exposed to them intensively.

Intentionality

Intentionality was introduced during Treatment 3 of the study. Prior to this, participants had performed tasks that had induced two incidental exposures to the 28 target collocations during Treatments 1 and 2. During Treatment 3, participants first performed a task that induced a third incidental exposure to 14 of the target collocations: the 3INC condition. Participants then studied the remaining 14 target collocations intentionally for an announced immediate posttest: the 2INC+INT condition. Three comparisons were made to test the learning effects of these two conditions.

Target Collocations Initially Learned

The first comparison focused on the target collocations that participants *knew* at Posttest 1 and whether the 3INC or the 2INC+INT condition was more effective in facilitating durable learning of them at the posttest that followed. For each participant two ratios from 0% to 100% were calculated based on the number of target collocations that participants knew at Posttest 1 and still knew at Posttest 2: one 3INC ratio and one 2INC+INT ratio (see Table 6).

Table 6
Ratios of Durability of Initially Learned Target Collocations

Learning condition	*N*	*M*	*SD*
3INC	45	71.47	33.581
2INC+INT	45	90.42	26.092

Note. N = number of participants, *M* = mean ratios in %, *SD* = standard deviations.

To find out whether the difference between the two means was statistically significant, a paired samples *t*-test was run. The result of this test was statistically significant: t (44): 3.100, p. = .003. This indicates that the introduction of intentional learning for the third exposure to target collocations that had been learned initially was more effective in promoting durable learning than a third incidental exposure to them.

Target Collocations Still Available for Learning

The second comparison focused on the target collocations that participants *did not know* at Posttest 1 and whether the 3INC or 2INC+INT condition was more effective in facilitating learning at the posttest that followed. For each participant two ratios from 0% to 100% were calculated based on the number of target collocations that participants did not know at Posttest 1 but knew at Posttest 2: one 3INC ratio and one 2INC+INT ratio (see Table 7).

Table 7
Ratios of Target Collocations Learned After Treatment 3

Learning condition	*N*	*M*	*SD*
3INC	45	30.09	28.87
2INC+INT	45	76.69	24.33

Note. N=number of participants, M=mean ratios in %, SD=standard deviations.

A paired samples *t*-test was run to find out whether the difference between the two means was statistically significant. The result of this test was significant: t (44): 10.059, $p. < 0.001$. This indicates that intentional learning of target collocations that had not been learned after the first exposure was more effective in promoting learning than a third incidental exposure.

Retention of Target Collocations

The third comparison focused on the results of the delayed posttest and whether the 3INC or 2INC+INT condition was more effective in facilitating lasting retention of the target collocations. The delayed posttest score was analyzed for each participant, and the number of target collocations that were learned in each of the two conditions was calculated (see Table 8).

Table 8
Delayed Posttest Scores

	Maximum score per condition: 14 points		
Learning condition	*N*	*M*	*SD*
3INC	45	6.89	3.43
2INC+INT	45	9.62	3.94

Note. *N* = number of participants, *M* = means, *SD* = standard deviations.

To find out whether the difference between the two means is statistically significant, a paired samples *t*-test was run. The result of this test was significant: t (44): 6.090, $p. < 0.001$. This indicates that the 2INC+INT condition was more effective in promoting lasting learning effects of target collocations than the 3INC condition.

Discussion

Research Questions

The first research question (RQ1) investigates whether Laufer and Hulstijn's (2001) Involvement Load Hypothesis (ILH) applies to the learning of English verb-noun collocations for adolescent L1 Swedish learners. Several methodological issues identified in previous ILH studies were considered when designing the present study to make it more robust, most notably the use of a within-subjects design. The answer to RQ1 is "no" when the effectiveness of the four tasks A–D was compared. The analysis showed that Task D, with the highest involvement load, was significantly less effective in promoting learning gains of target collocations than Tasks A–C, all with lower involvement loads. Furthermore, Task A was as effective as Tasks B and C, which it theoretically should not have been as it had a lower involvement load. These results run counter to the predictions of the ILH, but conform to previous studies that failed to find consistent significant positive learning effects of tasks with relatively higher involvement loads. One example is the study by Hulstijn and Laufer (2001), with Israeli and Dutch participants who performed three tasks inducing low (Task 1), middle (Task 2), and high (Task 3) involvement loads. While Task 3 outperformed Tasks 1 and 2 in both countries, Task 2 did not outperform Task 1 for the Dutch participants. It seems that the ILH is not fully capable of predicting the effectiveness of tasks on the learning of single L2 words and even less so for L2 collocations. One explanation for this result is that collocations are made up of several words that differ in their learnability as a function of their frequency and concreteness, and that it is therefore not possible to apply the ILH to lexical items larger than single words. Future research should investigate this issue with a more strict

control of target items, perhaps by using pseudowords as in Pellicer-Sánchez (2017) or technical terms as in Sonbul (2012).

The lack of support for the ILH in the present study may also be explained by lexical input processing (lex-IP) theory (Barcroft, 2015), which predicts that semantic elaboration of unknown words—for example, writing original sentences using them—facilitates the learning of new word *meaning* to the detriment of new word *form*. The reader is reminded that in Task D participants wrote original sentences using the target collocations after having searched for their meaning in a dictionary entry, thus performing two meaning-oriented activities. As participants were tested only on productive knowledge of form of target collocations (translation L1-L2), it may have disfavoured Task D. Lex-IP predicts that if a test of receptive knowledge of target collocation meaning had been administered, Task D would have proved most effective. Further support for lex-IP theory in relation to the present study is that Task B, in which participants also wrote original sentences using the target collocations but *did not* search for their meaning first, was more effective than Task D.

The second research question (RQ2) focuses on whether spacing effects influence learning of target collocations. The question was whether an expanding learning schedule would be more effective than an intensive learning schedule. The analysis showed that this was not the case, and the answer to RQ2 is therefore "no." This result is surprising considering the body of empirical support for the advantage of spreading out opportunities to learn new material rather than concentrating them. One interpretation of the result is that spacing effects in previous studies presupposed a comparison between purely *massed* learning schedules and spaced ones, not an *intensive* learning schedule as in the present study, and that this difference explains the result. Relatedly, Folse (2006) found that multiple retrievals[3] of target items was the strongest predictor of learning gains; it is possible that the intensive learning schedule induced such retrievals, which may have counterbalanced the positive effects of the expanding learning schedule and resulted in a lack of significant difference between the two schedules. Furthermore, the use of a between-subjects design for spacing effects in the study, which is less methodologically robust than within-subjects designs used for two other constructs under investigation, may also explain the result for RQ2. More research is needed to tease apart the effects of the variables in question using the more robust within-subjects designs.

The third research question (RQ3) centres on the effect of an intentional learning intervention with an announced posttest during the third exposure to target collocations (the 2INC+INT condition) compared to a third incidental exposure with no explicit forewarning of a posttest (the 3INC condition). The analysis showed statistically significant positive learning gains on all three measures for 2INC+INT: (a) the durability of target collocations that had been learned initially, (b) the learning of target collocations that had not

been learned after the first exposure, and (c) the retention—lasting effect—of target collocations as demonstrated in the delayed posttest. The answer to RQ3 is therefore "yes." This result implies that the benefits of intentional learning for single L2 words (cf. Elgort, 2011) also apply to collocations. Furthermore, it indicates that regardless of when the intentional intervention is introduced—before or after the initially form-meaning link has been established—asking learners to study target items intentionally is an effective approach compared to purely incidental learning conditions. The effect is lasting and with practical importance, as evidenced by the moderate effect size (0.7) when comparing the two conditions for the delayed posttest using Plonsky and Oswald's (2014) recalibrated interpretations of effect sizes for L2 research.

Pedagogical Implications

The most important pedagogical implication of the study concerns learning conditions of collocations as induced by the instructor: incidental vs. intentional. It seems that, in the absence of significant effects for various written tasks and intervals between exposures, the simple procedure of asking learners to study target items in a glossed list for an immediate posttest had a powerful effect on their ability to reproduce their form in an L1-L2 translation test. While this effect was unsurprising for the immediate posttest, it was surprising that it generated such a considerable return on investment in the delayed posttest. Another pedagogical implication of the findings is that original-sentence-writing cannot be recommended if instructors aim to help learners reproduce the form of new collocations. This recommendation aligns with the outcome of Folse's (2006) study that showed that multiple retrievals of single L2 words were more effective in promoting learning gains than original-sentence-writing. Folse elaborates on the drawbacks of original-sentence-writing for vocabulary learning (2006, p. 288) and recommends that instructors create sentences with gaps for target items and have learners choose from a set of options to fill in the gaps. This task was used in the present study for Tasks A and C for the test of the ILH, but its effectiveness was not substantiated (cf. the first subsection of the discussion on results, "The Involvement Load Hypothesis"). The most effective task for facilitating L2 collocation learning has not yet been found, but the best recommendation is still contrastive analysis and translation used by Laufer and Girsai (2008). These tasks involve learners' L1 in the learning process, which should be considered an asset and not a hindrance in the L2 classroom (see Schmitt, 2008, for discussion). The study also found that instructors need not consider spacing effects in their classroom practice, as spreading out exposures to target collocations did not significantly outperform an intensive learning schedule. Multiple retrievals of target items are required for stable entrenchment in the memory (Nation, 2001, pp. 66–67), but the intervals at which they occur do not seem to play a decisive role for L2 collocations.

Limitation of Study

One limitation of the study is the lack of control for the 28 target collocations. Even if they share several features, it may be that characteristics of the individual word components impacted on the learnability of the target collocations. This implies that it might have been easier to learn a target collocation with a high-frequency verb such as *carry* (in *carry a risk*), whose meaning it must be assumed that participants knew, than a low-frequency verb such as *rivet* (in *rivet the attention*), likely to be unknown to most participants. However, the novelty effect (Tulving & Kroll, 1995), which highlights the learnability of unknown material, provides counterevidence to this assumption. In addition, using target collocations with both high- and low-frequency word components can be argued to increase the ecological validity of the study, as learners under normal circumstances are likely to be exposed to FSs of various degrees of frequency (Webb et al., 2013, p. 93).

Conclusion

English language learners need to build up a large repertoire of verb-noun collocations to operate effectively in English. The most challenging type of such collocations are incongruent ones, and the most advanced level of knowledge is the ability to use them productively. This study found that the only effective way of helping learners accomplish this task was to ask them to study target collocations intentionally for an announced posttest. This result sends an important signal to instructors adhering to communicative approaches to language pedagogy that downplay the explicit/intentional teaching of vocabulary in favour of meaning- and production-oriented lesson activities. Meaning-focused instruction should by no means be abandoned in the language classroom, but it is high time that rote learning based on decontextualized glossed lists was re-evaluated in the language classroom because it is effective, efficient, and practical, for both single words and collocations.

Notes

1 Hulstijn and Laufer's (2001) study included six single words and four FSs, two of which were collocations (*morally derelict* and *deeply ingrained*), but not verb-noun collocations as in the present study.

2 Participants signed informed consent forms. The study had been approved by the Swedish Ethical Review Board (www.epn.se).

3 Retrieval occurs when a learner has established the form-meaning link of a word and then manages to do it again, cued by word form (receptive retrieval) or word meaning (productive retrieval) (Nation, 2001, p. 67).

Acknowledgements

I would like to thank Camilla Bardel, Batia Laufer, Hannah Neuser, and Tore Nilsson for their constructive comments on previous versions of this article.

The Author

Per Snoder has taught English, French, and Italian for 10 years in secondary schools in Sweden. He has also worked as a lecturer in Language Education at Stockholm University for four years. He is now a PhD student with a keen interest in formulaic language and language pedagogy.

References

Barcroft, J. (2015). *Lexical input processing and vocabulary learning*. Amsterdam, Netherlands: John Benjamins. https://doi.org/10.1075/lllt.43

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

Boers, F., & Lindstromberg, S. (2012). Experimental and intervention studies on formulaic sequences in a second language. *Annual Review of Applied Linguistics, 32*, 83–110. https://doi.org/10.1017/S0267190512000050

Boers, F., Demecheleer, M., Coxhead, A., & Webb, S. (2014). Gauging the effects of exercises on verb-noun collocations. *Language Teaching Research, 18*(1), 54–74. https://doi.org/10.1177/1362168813505389

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

Chan, T-P., & Liou, H-C. (2005). Effects of web-based concordancing instruction on EFL students' learning of verb-noun collocations. *Computer Assisted Language Learning, 18*(3), 231–250. https:// doi.org/10.1080/09588220500185769

Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment (CEFR).* Retrieved from http://www.coe.int/t/dg4/linguistic/Cadre1_en.asp

Doczi, B., & Kormos, J. (2016). *Longitudinal developments in vocabulary knowledge and lexical organization*. Oxford, UK: Oxford University Press.

Eckerth, J., & Tavakoli, P. (2012). The effects of word exposure frequency and elaboration of word processing on incidental L2 vocabulary acquisition through reading. *Language Teaching Research, 16*(2), 227–252. https://doi.org/10.1177/1362168811431377

Elgort, I. (2011). Deliberate learning and vocabulary acquisition in a second language. *Language Learning, 61*(2), 367–413. https://doi.org/10.1111/j.1467-9922.2010.00613.x

Eysenck, M. W. (1982). Incidental learning and orienting tasks. In C. R. Puff (Ed.), *Handbook of research methods in human memory and cognitition* (pp. 197–228). New York, NY: Academic Press.

Folse, K. S. (2006). The effect of type of written exercise on L2 vocabulary retention. *TESOL Quarterly, 40*(2), 273–293. https://doi.org/10.2307/40264523

Gyllstad, H. (2007). *Testing English collocations: Developing receptive tests for use with advanced Swedish learners*. Lund, Sweden: Lund University, Media-Tryck.

Henriksen, B. (2013). Research on L2 learners' collocational competence and development: A progress report. In C. Bardel, C. Lindqvist, & B. Laufer (Eds.), *Vocabulary acquisition, knowledge and use: New perspectives on assessment and corpus analysis*, Eurosla Monograph Series 2 (pp. 29–56).

Howarth, P. (1998). The phraseology of learners' academic writing. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis, and applications* (pp. 161–186). Oxford, UK: Clarendon Press.

Hulstijn, J. H. (2001). Intentional and incidental second language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 258–286). Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9781139524780

Hulstijn, J. H. (2003). Incidental and intentional learning. In C. Doughty & M. Long (Eds.), *The handbook of second language acquisition* (pp. 349–381). Oxford, UK: Blackwell.

Hulstijn, J. H., & Laufer, B. (2001). Some empirical evidence for the involvement load hypothesis in vocabulary acquisition. *Language Learning, 51*, 539–558. https://doi.org/10.1111/0023-8333.00164

Keating, G. (2008). Task effectiveness and word learning in a second language: The involvement load hypothesis on trial. *Language Teaching Research, 12*, 365–386. https://doi.org/10.1177/1362168808089922

Kim, Y. (2008). The role of task-induced involvement and learner proficiency in L2 vocabulary acquisition. *Language Learning, 58*(2), 285–325. https://doi.org/10.1111/j.1467-9922.2011.00644.x

Kornell, N. (2009). Optimizing learning using flashcards: Spacing is more effective than cramming. *Applied Cognitive Psychology, 23*(9), 1297–1317. https://doi.org/10.1002/acp.1537

Laufer, B., & Girsai, N. (2008). Form-focused instruction in second language vocabulary learning: A case for contrastive analysis and translation. *Applied Linguistics, 29*(4), 694–716. https://doi.org/10.1093/applin/amn018

Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning, 54*(3), 399–436. https://doi.org/10.1111/j.0023-8333.2004.00260.x

Laufer, B., & Hulstijn, J. H. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics, 22*, 1–26. https://doi.org/10.1093/applin/22.1.1

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

Lee, S., & Pulido, D. (2017). The impact of topic interest, L2 proficiency, and gender on EFL incidental vocabulary acquisition through reading. *Language Teaching Research, 21*(1), 118–135. https://doi.org/10.1177/1362168816637381

Lewis, M. (Ed.). (2000). *Teaching collocation: Further developments in the lexical approach*. Hove, UK: Language Teaching Publications.

Liu, D. (2010). Going beyond patterns: Involving cognitive analysis in the learning of collocations. *TESOL Quarterly, 44*(1), 4–30.

McEnery, T., & Xiao, R. (2011). What corpora can offer in language teaching and learning. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (Vol. 2, pp. 364–380). New York, NY: Routledge.

McIntosh, C., Francis, B., & Poole, R. (Eds.). (2009). *Oxford collocations dictionary for students of English*. Oxford, UK: Oxford University Press.

Meunier, F. (2012). Formulaic language and language teaching. *Annual Review of Applied Linguistics, 32*, 111–129. https://doi.org/10.1017/S0267190512000128

Nakata, T. (2011). Computer-assisted second language vocabulary learning in a paired-associate paradigm: A critical investigation of flashcard software. *Computer Assisted Language Learning, 24*, 17–38. https://doi.org/10.1080/09588221.2010.520675

Nakata, T. (2015). Effects of expanding and equal spacing on second language vocabulary learning. *Studies in Second Language Acquisition, 37*, 677–711. https://doi.org/10.1017/S0272263114000825

Nassaji, H., & Hu, H.-C. M. (2012). The relationship between task-induced involvement load and learning new words from context. *International Review of Applied Linguistics, 50*, 69–86. https://doi.org/10.1515/iral-2012-0003

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge, UK: Cambridge University Press.

Nesselhauf, N. (2005). *Collocations in a learner corpus*. Amsterdam, Netherlands: John Benjamins. https://doi.org/10.1075/scl.14

Pawley, A., & Syder, F. (1983). Two puzzles for linguistic theory: Native-like selection and native-like fluency. In J. Richards & R. Schmidt (Eds.), *Language and communication* (pp. 191–226). London, UK: Longman.

Pellicer-Sánchez, A. (2017). Learning L2 collocations incidentally from reading. *Language Teaching Research, 21*(3), 381–402. https://doi.org/10.1177/1362168815618428

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

Petti, V. (Ed.). (1994). *Norstedts comprehensive English-Swedish dictionary* (2nd ed.). Stockholm, Sweden: Norstedts.

Plonsky, L., & Oswald, F. L. (2014). How big is 'big'? Interpreting effect sizes in L2 research. *Language Learning, 64*(4), 878–912. https://doi.org/10.1111/lang.12079

Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language Teaching Research, 12*(3), 329–363. https://doi.org/10.1177/1362168808089921

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Basingstoke, UK: Palgrave Macmillan. https://doi.org/10.1057/9780230293977

Schmitt, N., & Carter, R. (2004). Formulaic sequences in action. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing, and use* (pp. 1–22). Amsterdam, Netherlands: John Benjamins. https://doi.org/10.1075/lllt.9

Schuetze, U. (2015). Spacing techniques in second language vocabulary acquisition: Short-term gains vs. long-term memory. *Language Teaching Research, 19*(1), 28–42. https://doi.org/10.1177/1362168814541726

Seabrook, R., Brown, G. D. A., & Solity, J. (2005). Distributed and massed practice: from laboratory to classroom. *Applied Cognitive Psychology, 19*(1), 107–122. https://doi.org/10.1002/acp.1066.

Sonbul, S. (2012). *The interface of explicit/implicit lexical knowledge: Acquisition and processing of collocations by natives and nonnatives* (Unpublished doctoral dissertation). University of Nottingham, England.

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*(2), 3–37.

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), 245–265. https://doi.org/10.1111/ijal.12092

Tulving, E., & Kroll, N. (1995). Novelty assessment in the brain and long-term memory encoding. *Pyschonomic Bulletin & Review, 2*(3), 387–390. https://doi.org/10.3758/BF03210977

Webb, S., & Kagimoto, E. (2011). Learning collocations: Do the number of collocates, position of the node word, and synonymy affect learning? *Applied Linguistics, 32*(3), 259–276. https://doi.org/10.1093/applin/amq051

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

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

Wu, S., Witten, I. H., & Franken, M. (2010). Utilizing lexical data from a Web-derived corpus to expand productive collocation knowledge. *ReCALL, 22*, 83–102. https://doi.org/10.1017/S0958344009990218

Ying, Y., & O'Neill, M. (2009). Collocation learning through an 'AWARE' approach: Learner perspectives and learning process. In A. Barfield & H. Gyllstad (Eds.), *Researching collocations in another language* (pp. 181–193). Basingstoke, UK: Palgrave Macmillan. https://doi.org/10.1057/9780230245327

Zou, D. (2017). Vocabulary acquisition through cloze exercises, sentence-writing and composition-writing: Extending the evaluation component of the involvement load hypothesis. *Language Teaching Research, 21*(1), 54–75. https://doi.org/10.1177/1362168816652418

Appendix A

Target collocations: pretest familiarity (the number of participants out of 44 who knew them productively and the corresponding percentages in brackets)

Target collocations	Pretest familiarity (raw frequencies and percentages)
1. approach problem	4/44 (9.1%)
2. attach importance	0/44 (0%)
3. bear child	3/44 (6.8%)
4. carry risk	0/44 (0%)
5. contract disease	1/44 (2.2%)
6. dent confidence	1/44 (2.2%)
7. entertain hope	0/44 (0%)
8. extend hospitality	0/44 (0%)
9. foot bill	0/44 (0%)
10. flag taxi	0/44 (0%)
11. harbour suspicions	1/44 (2.2%)
12. jog memory	2/44 (4.5%)
13. kick habit	3/44 (6.8%)
14. kindle interest	0/44 (0%)
15. level accusations	0/44 (0%)
16. pitch tent	3/44 (6.8%)
17. relax restrictions	0/44 (0%)
18. reap benefits	2/44 (4.5%)
19. rivet attention	0/44 (0%)
20. stir imagination	0/44 (0%)
21. strike balance	1/44 (2.2%)
22. shed clothes	3/44 (6.8%)
23. spell trouble	0/44 (0%)
24. sack employee	2/44 (4.5%)
25. shelve plan	1/44 (2.2%)
26. score success	1/44 (2.2%)
27. slash costs	0/44 (0%)
28. thumb ride	1/44 (2.2%)

Appendix B

For reasons of space, only one example of each type of instructional materials is provided below. All materials are available as online supplements. Appendix B1 contains one of the four ILH texts and the gap-fill task (A) that learners subsequently completed. Appendix B2 contains one of the six shorter texts that learners read and wrote a new title for (Task F).

Appendix B1

> I awoke this morning to a dusty village after pitching a tent last night in total darkness under a tree. It was freezing cold in the morning, but by 9 am the temperature was so high that I was forced to shed my clothes as my leather jacket and jeans were far too thick. I replaced them with lighter weight material. I knew that by traveling to Rwanda I would be running the risk of contracting a disease. After all, malaria is an ongoing epidemic here. At noon I started feeling dizzy and sick so I decided to go to the hospital. The village is very remote with no chance of flagging a taxi, but thankfully I was able to thumb a ride with a friendly local who further extended his hospitality by inviting me to dinner at his home. This was a pleasant contrast to the faint tension I've experienced from the local villagers who probably see me as an outsider and therefore harbour suspicions.

Task A

Use the expressions in the circle to complete the story

> "I awoke this morning to a dusty village after _____ last night in total darkness under a massive tree. It was freezing cold in the morning, but by 9 am the temperature was so high that I was forced to _____ as my leather jacket and jeans were far too thick and replace them with lighter weight material. I knew that by traveling to Rwanda I would be running the risk of _____. After all, malaria is an ongoing epidemic here. At noon I started feeling dizzy and sick so I decided to go to the hospital. The village is very remote with no chance of _____ but thankfully I was able to _____ with a friendly local who further _____ by offering me to have dinner at his home. This was a pleasant contrast to the faint tension in the air I've experienced from the local villagers who probably see me as an outsider and therefore _____."

thumb a ride – få lift

harbour suspicions– hysa misstankar

extend hospitality – visa gästfrihet

contract a disease – ådra sig (= få) en sjukdom

flag a taxi – hejda en taxi på gatan

shed the clothes – kasta av sig kläderna

pitch a tent – sätta upp ett tält

Appendix B2

Task F

Picking up strangers and driving them somewhere occurs in two ways: unpaid or paid. Hitch-hikers are a common sight outside petrol stations along highways in many countries. Equipped with a cardboard sign stating their destination, they put on a friendly face and entertain the hope of thumbing a ride to reduce the cost of travel and maybe have a conversation on the way. In big cities such free-riders are rare. City people instead turn to the street and flag a taxi if they are lost or to avoid being late for a meeting. Taxi-drivers need to find their way around but also to have social skills, which include striking a balance between interacting with customers and being too obtrusive. After all, some customers may want a moment of quiet and not hear the taxi-driver vent his anger about some political issue.