Applied Psycholinguistics (2022), **43**, 1–40
doi:10.1017/S0142716421000370

CAMBRIDGE
UNIVERSITY PRESS

ORIGINAL ARTICLE

Testing the reminding account of the lag effect in L2 vocabulary learning

Natalie G. Koval

Department of Linguistics and Languages, Michigan State University, East Lansing, MI, USA
Email: kovalnat@msu.edu

(Received 22 October 2020; revised 3 June 2021; accepted 24 June 2021; first published online 17 August 2021)

Abstract

Research has produced mixed findings regarding the effects of spacing L2 study. In order to know how this potentially very powerful learning tool can be useful, it is important to understand the cognitive mechanisms that drive the effects in L2 learning and how the operation of these mechanisms may be affected by variables relevant for SLA contexts. In this study, I examine the contribution of the dual mechanism of successful effortful retrieval during study to the lag effect in foreign vocabulary learning from L2-L1 retrieval practice. I additionally investigate the effects of feedback study time on the operation of the two cognitive mechanisms under investigation. Native speakers of English studied Finnish vocabulary during L2-L1 retrieval practice in paired-associate learning while their response latencies and accuracy were recorded. Results suggest that: (a) successful effortful retrieval underlies benefits of spacing L2-L1 retrieval practice: even with immediate feedback study, the benefits of effort are conditional on retrieval success; (b) successful retrieval is more beneficial than unsuccessful retrieval, contrary to proposals where this was not directly tested; and (c) imposing longer study time externally has little benefit, unlike what has been previously found with learner-regulated longer study time. Implications for L2 learning and teaching are discussed.

Keywords: the spacing and lag effects; L2 vocabulary acquisition; retrieval practice; reminding; study-phase retrieval

Learning large numbers of words is an important part of becoming proficient in a second language (L2). Therefore, an important question for L2 pedagogy is how to go about the task of learning and teaching vocabulary in a way that is both successful and efficient. L2 research has addressed this question by testing different methods of learning vocabulary. One method that has been widely found to enhance retention of studied material in the field of psychology is to space repeated study of target material (Crowder, 1976; Dellarosa & Bourne, 1985; Dempster, 1988, 1989; Hintzman, 1974; Pavlik & Anderson, 2005; Rohrer & Pashler, 2007; Wegener, Wang, Beyersmann, Nation, Colenbrander, & Castles, 2021). This finding, widely known as the spacing effect, has also been observed with learning of L2 vocabulary

(Bloom & Shuell, 1981; Koval, 2019; Nakata, 2015). A related finding, termed the *lag effect*, is the finding that how widely repeated study is spaced may have important consequences for learning outcomes (D'Agostino & DeRemer, 1973; Toppino, & Gracen, 1985).

The spacing effect is one of the most robust and ubiquitous findings in memory research. The benefits of spacing are usually very large: it is often found that two massed (consecutive) exposures to a target item are hardly more effective than a single exposure, while two spaced exposures are often twice as effective as one (e.g., Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006). The spacing effect potentially holds great promise for any learning situation. However, the full extent of its potential benefits is not being exploited in educational settings (Cepeda et al., 2009; Dempster, 1988; Gerbier & Toppino, 2015; Kang, 2016; Maddox, 2016). Further, investigations in the context of L2 acquisition have produced mixed results regarding spacing repeated study more widely, with some studies finding that this has either no effect or a detrimental effect on learning (Collins, Halter, Lightbown, & Spada, 1999; Elgort & Warren, 2014; Nakata, 2015; Nakata & Elgort, 2021; Rogers & Cheung, 2020a, 2020b; Serrano, 2011; Serrano & Muñoz, 2007; Suzuki & DeKeyser, 2017; White & Turner, 2005). In order to understand when and how spacing repeated study of L2 material may be beneficial and to be able to give useful practical recommendations regarding how to make the best use of this potentially very powerful learning tool in L2 pedagogy, it is important to understand the underlying cognitive mechanisms that drive the effects of spacing in L2 learning. It is further important to understand how the operation of these mechanisms may be affected by variables that are relevant for L2 learning contexts. Prior SLA research has tested the effects of spacing repeated study on the acquisition of various aspects of the second language and provides important information on the usefulness of this learning method for SLA contexts. However, this research has not produced much direct investigation into its underlying cognitive mechanisms and their interactions with relevant variables within the target learning contexts (although this criticism is not limited to the field of SLA, see, e.g., Dempster, 1988). The present study contributes to filling this gap. In the present study, I test the two-process mechanism of successful effortful retrieval during study as an underlying mechanism of spacing and lag effects in L2 vocabulary learning from retrieval practice in paired-associate learning (PAL). Such a dual mechanism is proposed to underlie spacing and lag effects within the *reminding* framework (Benjamin & Tullis, 2010). I further investigate how the operation of this mechanism may be affected by a variable that is relevant for second-language learning contexts, which is the amount of time a learner is given for studying a foreign word with its translation per encounter and in total (presented as feedback that follows each retrieval attempt), while holding the number of encounters constant. This latter variable is referred to, throughout this text, as feedback study time.

Explaining the spacing and lag effects

Despite the fact that research interest in the effects of spacing dates back over a century and despite the large number of theories that have been proposed in efforts to explain them (e.g., Benjamin & Tullis, 2010; Bjork & Allen, 1970; Challis, 1993; Chen, Paas, & Sweller, 2021; Dellarosa & Bourne, 1985; Estes, 1955; Glenberg, 1979; Greene, 1989; Jacoby, 1978; Küpper-Tetzel & Erdfelder, 2012; Landauer, 1969; Madigan, 1969; Melton,1970; Pavlik & Anderson, 2005; Raaijmakers, 2003; Rundus, 1971; Thios & D'Agostino, 1976; Zimmerman, 1975), their underlying mechanisms are still poorly understood (Kiliç, Hoyer, & Howard, 2013; Maddox et al., 2018). It is widely recognized today that a different mechanism, or combination of mechanisms, may underlie the effects of spacing in different learning situations or target tasks (Chen et al., 2021; Gerbier & Toppino, 2015; Glenberg & Smith, 1981; Greene, 1989; Kornell & Bjork, 2008; Russo & Mammarella, 2002). In the present study, I investigate the contribution of successful effortful retrieval (Pyc & Rawson, 2009), which is the dual mechanism proposed by the reminding account (Benjamin & Tullis, 2010), to the effects of spacing and lag in L2 vocabulary learning from L2-L1 retrieval practice.

The reminding account

The reminding account (Benjamin & Ross, 2010; Benjamin & Tullis, 2010; Hintzman, 2004; 2010; Tullis, Benjamin, & Ross, 2014) is currently a leading explanation for the lag and spacing effects. According to the reminding account, learning from repetition is beneficial when a repeated encounter with an item involves successful, but effortful retrieval of the information encoded at previous encounters. Thus, the reminding account is a dual mechanism account that combines the assumptions of the desirable difficulty and deficient processing proposals (Bjork, 1994, 1999; Jacoby, 1978), which hold that benefits of spacing come from the decreased effort or less attentional engagement that characterizes processing of an item when it repeats with only a very short interval between repetitions, with an important role for processing repeated events as repeated, or successfully retrieving previously encoded information during repeated study events (Hintzman, 2004, 2010; Thios & D'Agostino, 1976).

An important characteristic of the lag function (which is the function that relates lag to learning outcomes) is that it is nonmonotonic, or an inverted-U in shape (Cepeda et al., 2009; Cepeda et al., 2006; Cepeda, Vul, Rohrer, Wixted, & Pashler, 2008; Küpper-Tetzel & Erdfelder, 2012; Rohrer & Pashler, 2007). This means that increasing the inter-stimulus interval (ISI) is beneficial for learning but only to a point: very long ISIs may actually have negative effects on learning (Benjamin & Tullis, 2010; Cepeda, et al., 2006; Maddox, 2016; Peterson, Wampler, Kirkpatrick, & Saltzman, 1963; Young, 1971). In other words, there is a limit to how widely we can space repeated study before this begins to actually have a detrimental effect on learning outcomes. Because with increasing ISIs retrieval of a previous encounter requires more effort, which is beneficial for learning according to the reminding account, but retrieval is only likely to be successful within a limited range of ISIs, beyond which such retrieval may fail, resulting in detrimental effects on learning according to the reminding account, the two processes assumed by the reminding account can together explain the shape of this function. Figure 1 presents a rough conceptual illustration of changes in retrieval effort and success that may be expected with increasing ISIs. Here, we see that if we assume the two processes

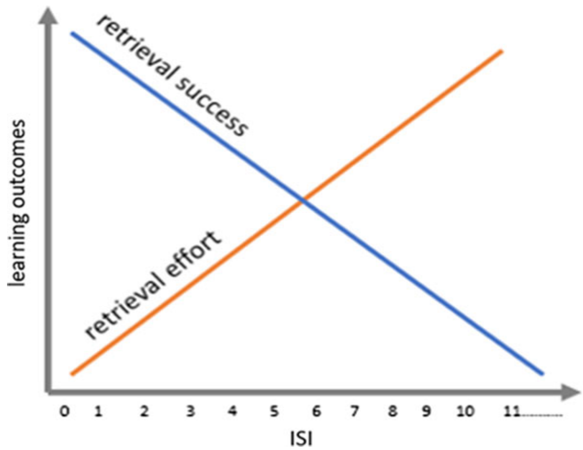

Figure 1. A conceptual illustration of changes in retrieval effort and success during training that can be expected with increasing ISI.

proposed by the reminding account, learning will be best at a point where both effort and success are at their highest and will be inferior where either one of these is low. Thus, the reminding account can explain the nonmonotonic shape of the lag function better than proposals that assume either of the two mechanisms as the sole underlying mechanism.

A number of findings that are potentially relevant for second-language learning can be accommodated if we assume an important role for successful retrieval during the study phase as it is assumed within the reminding account. One such finding is that optimal learning is at a higher level of ISI under intentional study than it is in incidental learning (Verkoeijen, Rikers, & Schmidt, 2005). Optimal learning here refers to the "sweet spot" represented conceptually by the point where the two lines intersect in Figure 1. This finding can be explained in terms of stronger memory traces that are laid down during intentional study, which can survive longer ISIs. Another important finding is a detrimental effect of spacing on learning in situations where repeated exposures occur in different rather than similar contexts (Verkoeijen, Rikers, & Schmidt, 2004). Further, study time has been found to positively affect learning from spaced repetitions (Verkoeijen & Bouwmeester, 2008), while task complexity and the difficulty of the intervening activity coupled with lower working memory capacity have been shown to negatively affect learning from spaced repetitions (Bui, Maddox, & Balota, 2013; Donovan & Radosevich, 1999). Thus, the findings that positive effects of spaced study may be tempered or even reversed under certain levels of the relevant variables can be explained through this affecting the probability of retrieval success during study (see, also, Suzuki, Nakata, & Dekeyser, 2019).

The focus of the present investigation on a dual-process account that includes successful retrieval during study as an underlying mechanism is motivated by the fact that a failure to process repeated encounters with target items as repetitions has been cited, though not directly tested, in SLA research as a potential explanation for failures to observe benefits of spacing (see, e.g., Elgort & Warren, 2014; Serrano, 2011). The inclusion of the second element of effortful processing is motivated by the widely held belief that attentional engagement and effort are beneficial for learning of second-language vocabulary (Godfroid, Boers, & Housen, 2013; Laufer & Hulstijn, 2001; Schmitt, 2008) as well as the finding that deficient processing of massed encounters mediates the benefits of spacing in L2 vocabulary learning (Koval, 2019).

Study time

There is potentially a large number of variables in L2 learning contexts that may affect the operation of the mechanisms of retrieval effort and success. The present study tests the moderating effects of one such variable, which is externally predetermined feedback study time. Longer study time might promote retrieval success with spaced repetitions due to stronger encodings at each repetition that are more likely to survive longer lags (Verkoeijen & Bouwmeester, 2008). Verkoeijen and Bouwmeester inferred such an underlying process from their finding that their lower-performing group benefitted from spaced practice only under the longer study time condition. They did not, however, directly test this possibility. Further, intuitively, longer study time might also reduce retrieval effort. Thus, study time might moderate the effects of spacing L2-L1 retrieval practice on the underlying mechanisms of retrieval effort and success and thus affect learning.

Longer study time might also have an independent effect on learning. In both psychology and SLA, studies show that the more time a learner spends studying a word, the better the learning outcomes are (Godfroid et al., 2018; Godfroid, et al., 2013; Koval, 2019; Rundus, 1971). Importantly, longer study time has also been shown to mediate the benefits of spacing in L2 word learning (Koval, 2019). The latter finding suggests that increased attentional processing underlies the benefits of spacing L2 vocabulary study. Findings from studies investigating self-regulated study time allocation suggest that learners tend to overestimate their knowledge of items in massed practice and devote less study time to these (Benjamin, Bjork, & Schwartz, 1998; Kornell & Bjork, 2007; Koval, 2019; Rundus, 1971; Shaughnessy, Zimmerman, & Underwood, 1972; Zechmeister & Shaughnessy, 1980; Zimmerman, 1975). Generally, learners are known to be quite ineffective at pacing their own study (Benjamin et al., 1998; Jacoby, Bjork, & Kelley, 1994; Kornell & Bjork, 2007). An interesting question that has important practical implications is whether externally predetermined longer study time affects learning in the same way as does learner-regulated longer study. If learners tend not to be effective at pacing their study, can we enhance learning by controlling the pace at which words are studied?

Retrieval practice

In this study, participants learn vocabulary from L2-L1 retrieval practice. Retrieval practice has been shown to enhance learning of studied material, including learning of L2 vocabulary (Barcroft, 2007; Carrier & Pashler, 1992; Cull, Shaughnessy, & Zechmeister, 1996; Karpicke & Roediger, 2008; Nakata, 2015, 2016; van den

Broek, Takashima, Segers, & Verhoeven, 2018). The act of retrieval has been shown to slow and otherwise interfere with forgetting of learned information (Hogan & Kintsch, 1971; Izawa, 1970; Maddox & Balota, 2015; Runquist, 1986; Wheeler & Roediger, 1992). Retrieval practice may further often constitute more transfer-appropriate processing for many skills (Kolers & Roediger, 1984; McDaniel, Friedman, & Bourne, 1978), such as when the meaning of an L2 word must be retrieved during comprehension of L2 input. Retrieval practice is believed to be more beneficial the more effortful, or complete the retrieval (Bjork, 1975; Glover, 1989; Pyc & Rawson, 2009; Whitten & Bjork, 1977). In fact, even when increased effort means more retrieval failures or errors during the learning phase, this is still argued to result in better retention in the long term (Pashler, Zarow, & Triplett, 2003; Schmidt & Bjork, 1992; Soderstrom, Kerr, & Bjork, 2016; Storm, Bjork, & Storm, 2010). Unsuccessful retrieval attempts are still known as powerful learning events because they are believed to promote deeper processing of the feedback that follows (Arnold & McDermott, 2013; Hays, Kornell & Bjork, 2013; Izawa, 1970; Kornell, Hays, & Bjork, 2009; Roediger & Karpicke, 2006a).

Just as is the case with the spacing effect, retrieval practice has been widely found to be beneficial for learning. Just as is the case with the spacing effect, however, its full potential has not been used in education (McDaniel & Fisher, 1991; Roediger & Karpicke, 2006b). Given that retrieval practice improves learning and that repeated retrieval further enhances learning (Karpicke & Roediger, 2008), an important question is what role the temporal distribution of such practice may play (Nakata, Tada, Mclean, & Kim, 2021). Spaced retrieval practice combines the benefits of spacing and retrieval and thus potentially maximizes learning. How best to use it is still a question, however (Storm et al., 2010).

Retrieval practice and spacing effects in SLA

Effects of spaced practice have received some attention in the field of SLA (Bird, 2010; Bloom & Shuell, 1981; Kasprowicz, Marsden, & Sephton, 2019; Lee, Maechtle, & Hu, 2021; Miles, 2014; Miles & Kwon, 2008; Nakata, 2015; Nakata & Suzuki, 2019; Nakata & Webb, 2016; Rogers, 2015; Rogers & Cheung, 2020a, 2020b; Schuetze, 2015; Serrano & Huang, 2018, 2021; Suzuki, 2017; Suzuki & DeKeyser, 2017; Suzuki & Sunada, 2020). While many of these studies have found benefits of spacing L2 study, others have found no effect or even a detrimental effect of spacing repetitions more widely (Collins et al., 1999; Elgort & Warren, 2014; Nakata, 2015; Rogers & Cheung, 2020a, 2020b; Serrano, 2011; Serrano & Huang, 2021; Serrano & Muñoz, 2007; Suzuki & DeKeyser, 2017; White & Turner, 2005) or that the finding of such an effect may depend on the learning outcome measure used (Nakata & Elgort, 2021). In order to know when and how spacing practice may be useful for L2 learning, it is important to understand the cognitive mechanisms that underlie effects of spacing in our specific learning situations and how the operation of these cognitive mechanisms may be affected by variables inherent in L2 learning contexts. SLA research has, thus far, focused mainly on the question of whether or not spacing affects the acquisition of different aspects of a second language, without much focus on the process as well as the product of learning. There are a few exceptions, however (Koval, 2019; Nakata & Suzuki, 2019; Suzuki, Nakata, & Dekeyser, 2019, 2020). Nakata and Suzuki (2019), for instance, measured learners' retrieval success during the study phase through the task of overt L2-L1 translation. However, the authors broke down the study of their 48 target words into 2 sets of 24 to minimize retrieval failure during study and did not test its effects on learning. Suzuki and DeKeyser (2017) included an ad hoc analysis of lexical retrieval performance during training on an element of L2 Japanese morphology and speculated that ease and success of lexical retrieval may affect the nature of cognitive processes involved in distributed and massed learning. Another study that has investigated the process as well as the product of learning under differential spacing is Koval (2019). In Koval (2019), I used eye-tracking methodology to measure reading times and showed that diminished processing of L2 words studied in a massed fashion during sentence reading mediated the large benefits of spacing obtained in my study.

More research is needed that explores the process as well as the product of learning L2 material under different levels of ISI. The present study contributes to an understanding of this process by examining the effects of successful effortful retrieval during study. Both retrieval success and effort may depend on a number of factors. One such factor may be the amount of time a learner is given for studying an L2 word with its translation per repetition. The longer the study time, the stronger the resulting encodings are likely to be (Verkoeijen & Bouwmeester, 2008), resulting in a higher level of retrieval success. At the same time, stronger encodings may mean less retrieval effort during subsequent presentations. Thus, feedback study time may have important consequences for the operation of both underlying mechanisms investigated in the present study.

Retrieval practice has been shown to be beneficial for L2 vocabulary learning (Barcroft, 2007; van den Broek et al., 2018). Further, psychology studies using L2 words as the learning targets have generally obtained benefits of spaced retrieval practice over massed retrieval practice as well as benefits of retrieval over restudying, particularly when knowledge is tested after a longer retention interval (RI), such as on delayed tests (Arnold & McDermott, 2013; Bahrick et al.,1993; Carrier & Pashler, 1992; Karpicke & Roediger, 2008; Pashler et al., 2003; Pavlik & Anderson, 2005; Pyc & Rawson, 2009). A similar finding by Nakata et al. (2021) is that cumulative L2 vocabulary tests are more beneficial than noncumulative tests. The authors suggest that one of the reasons cumulative tests may produce learning benefits is that learners are forced to review and retrieve the target information in a distributed fashion.

What is still not clear, however, is whether retrieval success is important, as proposed in the reminding account, when retrieval attempts are followed by feedback. It has been argued that, when feedback is provided, failure to retrieve L2 vocabulary information is beneficial for learning outcomes (Bahrick & Hall, 2005; Pashler et al., 2003). However, in these studies, effects of retrieval failure were not directly tested but only inferred. The present study directly tests the effects of retrieval failure during study that might result from longer ISI on learning outcomes. The present study investigates three levels of ISI within a declarative knowledge acquisition task. During study, retrieval effort should be highest at the longest ISI and lowest at the shortest ISI, while retrieval success should show the opposite pattern. If the dual mechanism proposed by the reminding account underlies the benefits of spacing L2 vocabulary study more widely, retrieval effort and success should together mediate the benefits of longer ISI. Further, the feedback study time variable should affect both retrieval success and effort during training, whereby words that are studied under the longer study time condition should be retrieved more successfully but also with less effort during each subsequent repetition.

Method

The present study is motivated by the following research questions:

1. Does the amount of lag between repeated retrieval-restudy events affect learning of L2 vocabulary?
2. Does feedback study time affect the relationship between lag and learning of L2 vocabulary?
3. Does successful effortful retrieval mediate the effects of lag on L2 vocabulary learning?
4. Does feedback study time affect the operation of the two mechanisms of retrieval success and effort during study?

Participants

Fifty-two native speakers of English (young adults) participated in the experiment. These were mostly undergraduate students in a wide variety of majors at Michigan State University who had responded to an ad placed through the Office of the Registrar. Twenty-two were male and 30 were female, aged 18–29 years ($M = 20.04$, $SD = 2.08$, Median = 20). Most of these students had studied at least one foreign language. None reported being familiar with the Finnish language. The participant sample size was based on previous research that has successfully used a similar population and materials to ask similar research questions (Koval, 2019). Throughout the results section, I will discuss the informational value of the results given the present sample size. All 52 participants completed the experiment with the exception of 4 participants who did not return for the delayed posttests and thus did not provide delayed posttest data. Based on a suggestion from two anonymous reviewers, these four participants' data were excluded from the analyses.

Materials and design

A fully counterbalanced within-item within-participant design was used. The experiment consisted of a study phase, a distractor math task, 30-min delayed vocabulary posttests (referred to as immediate posttests), and 1- to 2-week delayed vocabulary posttests (referred to as delayed posttests).

Study phase

Finnish was selected as the target language for the study. Finnish is a relatively uncommon L2 for US students. Being a language of the Finnic family, it also bears little resemblance to English or languages that are commonly studied by US students. Further, Finnish is written in the same alphabet as English, the participants' L1, which allowed to control for reading difficulty. All diacritic marks were removed from all Finnish words used in the experiment.

Seventy-two Finnish nouns were selected as the target words. None of these nouns were cognates of their English translations. The 72 target words were divided into 2 main lists (36 words each). The words on each list served as repeated targets half of the time and as once-presented controls the other half. The purpose of the unrepeated controls was to investigate the effects of retrieval practice in the three ISI conditions against a baseline of no retrieval practice. Within the lists, the words were further divided into three ISI sublists (12 words each), each to be used in each of the three levels of ISI (massed, short-spaced, and long-spaced) when serving as repeated targets. Each ISI list was further divided in half for the two levels of feedback study time (3 vs. 9 s). The feedback study time variable was operationalized as the duration of time feedback in the form of the target word pair stayed on the screen. Study-phase instructions asked the participants to study and rehearse the feedback for as long as it remained on the screen and to continue doing so even if they felt that they had learned a given word pair sufficiently. Thus, this variable was used to explore the effects of active engagement with the study of the word pairs for different durations of time. The two levels of feedback study time were determined as short and long based on the presentation duration used in other similar studies (e.g., Nakata & Suzuki, 2019) as well as participants' comments during the piloting stage. The two study time lists were matched on the number of letters. Four to five participants fell into each of the 12 lists that resulted from full counterbalancing.

The target words ranged in length from four to eight letters (see Appendix A). The N-Watch program (Davis, 2005) was used for information on frequency of the English translations. CELEX frequency and LOG 10 frequency were used. In N-Watch, LOG 10 frequency is based on the CELEX English Linguistic Database (Baayen, Piepenbrock, & van Rijn, 1995). Brysbaert, Warriner, and Kuperman's (2014) database of concreteness ratings was used for indices of concreteness. Appendix B presents frequency and concreteness information for the English translations in each condition. The target nouns were matched exactly on the number of letters between all lists and sublists. Two hundred and ten additional Finnish words were selected to serve as practice and recency items as well as filler trials during the study phase. The fillers were used to achieve the desired order of target items as well as to increase orthographic interference. They also served to prevent participants from noticing a pattern of repetition among the target words. The filler items were similar to the target items in form. Care was taken to exclude potential study-phase fillers that stood out as overly similar in form to the target items, the posttest distractor items, or to each other. Some of the fillers were followed by translations and others were not. Some fillers repeated and others were only presented once.

The target words repeated six times. The study phase was divided into six experimental blocks with 6-min breaks in between. This was done to allow participants breaks as well as to organize the distribution of the six repetitions of a word more neatly. The words in the massed condition repeated six times within each block. These were separated by 0–1 intervening trials. While massed practice, in its strictest

https://doi.org/10.1017/S0142716421000370 Published online by Cambridge University Press

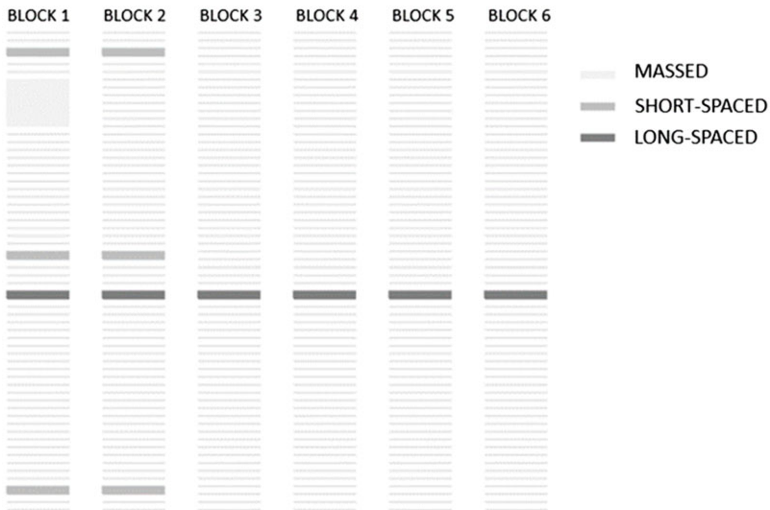

Figure 2. A conceptual illustration of the repetition pattern for one item in each condition.

sense, involves no intervening trials between repetitions (e.g., Cepeda et al., 2006), intervening trials were included here in order to dilute the six consecutive repetitions, particularly with regard to how predictable each next trial was as a repetition of the previous trial. However, the intervening Finnish words that separated repetitions in the massed condition were always fillers that were not accompanied by a translation in order to preserve the massed nature of study of word meanings. Thus, study in the massed condition was never interrupted by study of another form-meaning pair. The words in the short-spaced condition repeated over two consecutive blocks (three times per block) and were separated by 17–38 trials within a block and by 12–22 trials plus the 6-min distractor math task between two adjacent blocks. The words in the long-spaced condition repeated once per block and were separated by 71–119 trials plus the 6-min intervening distractor math task. The once-presented control words were distributed more or less evenly throughout the study phase. The average position across the experimental sequence was equated for the words in all four conditions (massed: 249.82; short-spaced: 249.97; long-spaced: 251.10; and controls: 248.44). Figure 2 presents the conceptual pattern of repetition for one item in each ISI condition across the six blocks. ISI was a within-participant manipulation, which means that each participant studied words under all three ISI conditions.

Each experimental block started and ended with three filler items. The conditions were equally represented at the beginnings and ends of blocks: blocks 2, 4, and 6 began and ended with two control items; block 1 began and ended with an item from the massed condition (all six repetitions); block 3 began and ended with two items from the short-spaced condition (1 repetition); and block 5 began and ended with two items from the long-spaced condition (1 repetition).

A practice block preceded the experimental sequence. A recency block followed the sixth experimental block. These blocks contained some of the same fillers that were used in the study phase. The purpose of the recency block was to minimize any recency or order effects on the 30-min delayed (immediate) posttest. All fillers that were translated were used in the practice and recency blocks. To resemble target items, these had to be translated. The rest of the fillers were not translated in order not to overwhelm participants with the number of translations they had to memorize. Fillers that were associated with their L1 translations were not in any way different from the target words from the point of view of the participant. Further, these often repeated in a similar pattern to the target words, except that the number of repetitions and the pattern of repetition was different and more haphazard. This was done to prevent participants from anticipating a pattern of repetition for the target items. The practice block served to minimize any effects of primacy as well as to familiarize the participants with the procedure.

Distractor math task

Participants performed addition, subtraction, multiplication, and division operations. Sometimes, they were asked to do mental math; other times, they were allowed to use paper. Such variety in activity was used to minimize boredom and fatigue.

Posttests

Three sets of paper-and-pencil posttests (see Appendix C) were used and were administered in a fixed order. Posttest 1 was a form recognition test. Here, the 72 target words were presented among 156 new Finnish words (distractors) that had not occurred during the study phase. Participants were to underline words that they recognized as ones studied during the study phase.

Posttest 2 was an L2-L1 translation test. Here, participants were to write the English translations next to the target Finnish words (on Sheet A).

Posttest 3 was a form-meaning matching test. Here, participants were presented with the English translations for all the target Finnish words (Sheet B). Participants were to write the numbers associated with English translations on Sheet B next to the corresponding Finnish words on Sheet A, which had been used in Posttest 2. Identical tests were used in immediate and delayed administrations except for order randomization between and within participants. A different set of distractors was used for the immediate and delayed form recognition tests, however.

Background questionnaire

Information was collected on participants' age, sex, languages studied, and any other information that the participants felt was relevant. The questionnaire also asked to indicate whether any of the studied words had struck participants as familiar upon initial encounter and to elaborate if the answer was yes.

Instruments

The DMDX software (Forster & Forster, 2003) was used on an *HP* laptop computer for stimulus presentation and recording of response latencies. Two *Transcend* voice recorders were used to record participants' oral responses.

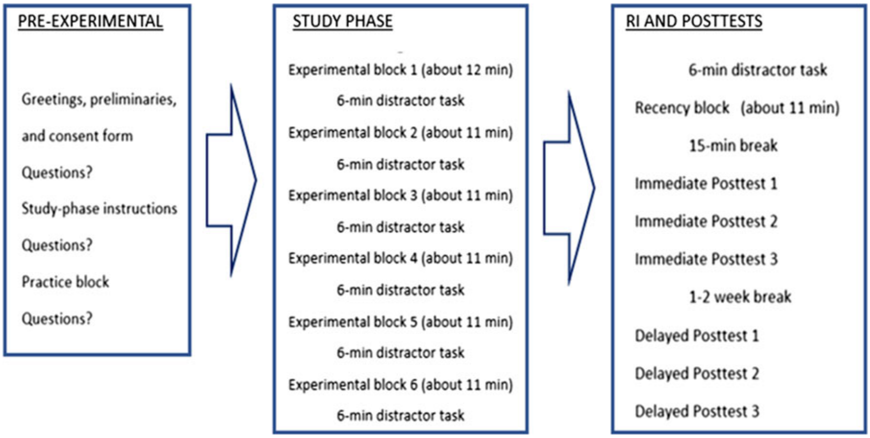

Figure 3. A summary of the experimental procedure.

Procedure

The experimental procedure is summarized in Figure 3. The entire experiment was approximately 3 hr and 45 min in duration, over two sessions, per participant. Session one was about 3 hr and 10 min. Session two was between 20 and 35 min. Session one included the study phase, a 15-min break, the immediate posttests, and the background questionnaire. Session two included only the delayed posttests. The two sessions were separated, depending on participant availability, by approximately 1 or 2 weeks. The experiment was conducted with each participant individually in a small quiet lab.

The practice block consisted of 83 trials. After and during the practice block, the participants were encouraged to ask any questions they might have. Following the completion of the practice block, the experimental blocks were completed in order, separated by 6-min distractor tasks. Block one consisted of 110 trials. Each subsequent block consisted of 90 trials. Block one took 12 min, on average, and each subsequent block took 11–12 min, on average, to complete.

Figure 4 presents an example of an experimental study-phase trial sequence. Each trial started with a row of hash symbols presented in the center of the screen for 1 s, after which it was replaced by a Finnish word prompting participants to produce its English translation. The participants were to say the translations aloud as quickly and accurately as they could while their responses were audio-recorded. If a participant could not remember a translation or if they thought that they had never seen the translation for a given word, they were to say "I don't know." Response time was recorded through a button press by the researcher, which initiated the next screen, on which the L1 translation appeared opposite the Finnish word. The pair stayed on the screen for 3 s or 9 s, depending on the level of exposure duration assigned to the word for the specific rotation version, after which the next trial began. Participants had been instructed to study each word pair for as long as it remained on the screen and not to stop studying it even if they felt that it had been sufficiently learned. Distractor words that were presented with translations followed the same sequence.

https://doi.org/10.1017/S0142716421000370 Published online by Cambridge University Press

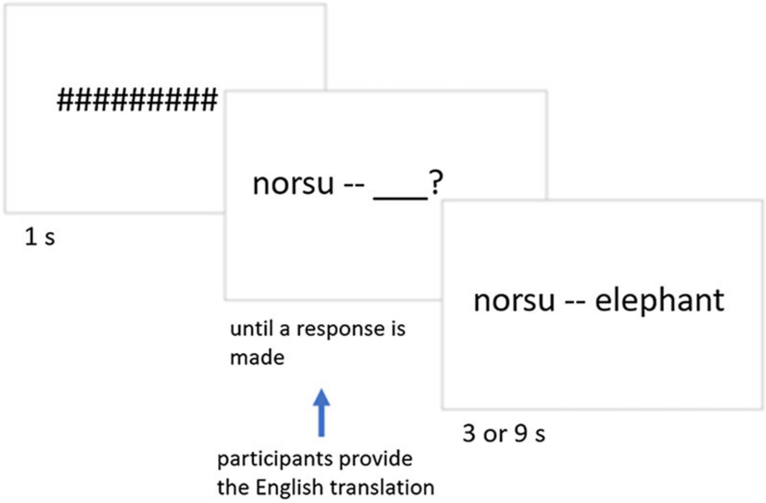

Figure 4. An example of an experimental study-phase trial sequence.

If a distractor word was not presented with a translation, the button press initiated the next trial. The researcher asked the participants how they were feeling at the end of each block, to which all responded that they were feeling good. However, based on the observation, during piloting, that many participants felt like it was difficult to remain seated for the entire duration of the study phase, after blocks 4, 5, and 6, the researcher suggested a walk outside the lab as part of the distractor math task. During the walk, participants performed mental math operations that the researcher asked them to perform. A few participants indicated that they did not feel like taking a walk – these participants performed the distractor math task in its entirety in the lab. The math task was mostly interactive, which was also done to cut down on possible fatigue.

After the recency block, there was a 15-min break, during which participants were free to leave the lab. After the break, participants performed posttests 1, 2, and 3, in order, untimed, and completed the background questionnaire. Participants were asked to return for the second session 2 weeks after session one. However, not all participants were able to come back in exactly 2 weeks. For those who were not, session two was mostly conducted with a shorter interval between the two sessions. Participants can be divided into two groups: 21 participants who came back 6–8 days after session one and 26 participants who came back 11–16 days after session one. Following the suggestion of one of the anonymous reviewers, I additionally performed analyses only on the data from participants who came back after a 2-week period. This did not change the pattern of results.

Participants were not told anything about the content of the second session. Session two was identical in content to the immediate posttests. At the end of session two, participants were asked whether they had had any exposure to the target Finnish words outside of the lab between the two sessions. This was noted by the researcher. All participants except one (whose delayed posttest data were removed from the analysis) stated that they had had no such exposure.

https://doi.org/10.1017/S0142716421000370 Published online by Cambridge University Press

Analyses and results

SPSS version 25 (IBM Corp., 2017) was used for all statistical analyses. Microsoft Office 365 Excel and PowerPoint were used for data management and some of the graphics. Linear mixed modeling and moderated mediation analyses were used. All statistical analyses were two-tailed and conducted at an alpha level of .05 except for cases where a Bonferroni correction was performed. Cohen's d effect sizes were calculated for the study phase and posttest results and interpreted in terms of the benchmarks suggested by Plonsky & Oswald (2014) but also in terms of the comparisons being made, which can also affect the substantive interpretation of an effect size in important ways. The posttests were scored as follows: one point was awarded for each correct response and zero points were awarded for an incorrect response or no response. On the translation test, synonyms that were very close to the target meanings (e.g., *cigar* for *cigarette* – there were no synonyms that did not have a very close meaning to the target word) as well as slight misspellings (there were no serious misspellings) were counted as correct responses.

Background questionnaire

See the *Participants* section for the demographic information collected through the background questionnaire. Six participants noted that some or many of the words looked like Spanish words or words from other languages in terms of the spelling. The rest of the participants indicated that none of the words had struck them as familiar or elaborated on their mnemonic devices, such as breaking words down into "mini words" that helped to "remember the whole word."

Posttest results

To answer the first and second research questions, posttest results were examined as a function of ISI and feedback study time. Reliability (Cronbach's α) for the posttests was as follows: immediate form recognition: $\alpha = .694$; immediate L2-L1 translation: $\alpha = .790$; immediate form-meaning mapping: $\alpha = .789$; delayed form recognition: $\alpha = .779$; delayed L2-L1 translation: $\alpha = .724$; delayed form-meaning mapping: $\alpha = .882$. According to Plonsky and Derrick (2016), this reflects medium to moderately high reliability in comparison to the reliability that has been reported for other instruments in the field. Form recognition accuracy was acceptable for all participants ($< 10\%$ error) except for two participants on the immediate test and one participant on the delayed test. These participants' data were excluded for the corresponding tests.

Posttest results: Descriptive statistics

Table 1 presents percent correct scores on the immediate and delayed posttests separately in the experimental and control conditions. The table shows that there is a positive effect of practice on all the tests, across the two test administrations. There is further a small positive overall effect of longer feedback study time in both conditions across the test types, in both test administrations.

https://doi.org/10.1017/S0142716421000370 Published online by Cambridge University Press

Table 1. Percent correct in the practice and no-practice conditions

	Experimental repeated items								Control items							
	Short presentation duration				Long presentation duration				Short presentation duration				Long presentation duration			
	M	N	SD	Mdn	M	N	SD	Mdn	M	N	SD	Mdn	M	N	SD	Mdn
Immediate posttests																
Form recognition	64.5	46	16.0	69.4	64.3	46	17.9	66.7	12.0	46	9.4	11.1	13.3	46	10.6	11.1
L2-L1 translation	53.6	48	20.6	55.6	58.9	48	17.9	61.1	3.0	48	4.7	0.0	6.2	48	7.0	5.6
Form-meaning matching	65.9	48	17.6	72.2	70.9	48	13.5	72.2	10.8	48	10.8	8.3	14.9	48	12.3	11.1
Delayed posttests																
Form recognition	53.8	47	17.9	50.0	56.4	47	17.0	61.1	13.7	47	11.1	11.1	16.1	47	13.0	16.7
L2-L1 translation	27.7	48	15.4	27.8	32.4	48	16.8	33.3	2.0	48	3.9	0.0	2.3	48	3.6	0.0
Form-meaning matching	39.1	48	18.8	36.1	44.0	48	19.7	44.4	3.8	48	7.0	0.0	4.6	48	7.7	0.0

Table 2. Percent correct in the three ISI conditions

	Massed practice				Short-spaced practice				Long-spaced practice			
	M	N	SD	Mdn	M	N	SD	Mdn	M	N	SD	Mdn
Short presentation duration												
Immediate tests												
Form recognition	32.6	46	21.1	33.3	78.6	46	21.6	83.3	82.2	46	18.1	83.3
Translation	16.3	48	17.0	16.7	72.6	48	27.0	83.3	71.9	48	28.6	83.3
Form-meaning matching	25.3	48	22.0	16.7	85.4	48	20.8	100.0	86.8	48	21.5	100.0
Delayed tests												
Form recognition	28.4	47	22.0	33.3	66.3	47	24.9	66.7	66.7	47	25.1	66.7
Translation	6.6	48	10.7	0.0	35.4	48	24.0	33.3	41.0	48	22.8	50.0
Form-meaning matching	10.8	48	13.5	0.0	50.7	48	28.6	50.0	55.9	48	22.9	50.0
Long presentation duration												
Immediate tests												
Form recognition	33.3	46	21.9	33.3	79.7	46	23.5	83.3	79.7	46	22.5	83.3
Translation	19.1	48	19.1	16.7	78.8	48	24.7	83.3	78.8	48	23.8	83.3
Form-meaning matching	33.7	48	22.7	33.3	90.6	48	15.7	100.0	88.5	48	17.9	100.0
Delayed tests												
Form recognition	28.0	47	19.7	33.3	70.6	47	22.8	83.3	70.6	47	22.6	66.7
Translation	7.6	48	11.9	0.0	39.6	48	24.9	33.3	50.0	48	28.6	50.0
Form-meaning matching	12.2	48	15.3	0.0	54.5	48	29.9	58.3	65.3	48	27.7	66.7

Table 2 presents percent correct scores in the three ISI conditions separately. Across the tests, there is a beneficial effect of spacing practice. In fact, median scores on both delayed meaning tests in the massed condition are zero, across the two levels of feedback study time. Although increasing the time a learner spends studying feedback per repetition and in total appears to benefit learning, spacing practice appears to have a much greater benefit, based on these percent correct descriptive statistics. The benefits of lag, however, do not appear to be as large or consistent: the long-spaced condition has produced scores that are only a bit larger than those in the short-spaced condition and this difference looks to be mostly limited to the delayed scores. Separate analyses further showed that the overall pattern of descriptive statistics was similar between the two genders.

Posttest results: Inferential statistics

ISI, RI (immediate vs. delayed test), and feedback study time were included as the independent variables in an omnibus test for each test type. Percent correct was used as the dependent variable. Due to high collinearity between the two variables of ISI and the variable that distinguishes experimental items from control items, these were collapsed into one variable that will be referred to as practice type. Because participants varied in the time between study-phase and delayed posttests, a random slope was included for the RI variable. Simultaneous entry with restricted maximum likelihood estimation was used.

The residuals for the form recognition test were close to normally distributed with two outliers beyond −3SD and two outliers beyond 3SD, which were removed. This resulted in a normal distribution according to the Kolmogorov–Smirnov ($p = .200$) and Shapiro–Wilk ($p = .153$) tests of normality. The distribution further had skewness and kurtosis within acceptable ranges (skewness = −.164, $SE_{\text{skewness}} = .090$; kurtosis = −.1.09, $SE_{\text{kurtosis}} = .179$). The ICC for the effect of participant was .055. The residuals for the L2-L1 translation test were close to normally distributed with two outliers beyond −3SD. After the removal of these outliers, the distribution was normal according to the Kolmogorov–Smirnov ($p = .200$) and Shapiro–Wilk ($p = .203$) tests of normality. The distribution further had skewness and kurtosis within acceptable ranges (skewness = −.159, $SE_{\text{skewness}} = .088$; kurtosis = −.100, $SE_{\text{kurtosis}} = .176$). The ICC for the effect of participant was .041. The distribution of the residuals for the form-meaning matching test was close to normally distributed with five outliers beyond −3SD. These outliers were removed, which resulted in a more nearly normal distribution (skewness = −.267, $SE_{\text{skewness}} = .089$; kurtosis = .260, $SE_{\text{kurtosis}} = .177$). The Kolmogorov–Smirnov and Shapiro–Wilk tests of normality were not significant at the .001 alpha level, ($p = .036$ and .002, respectively). Further, the distribution looked symmetrical and bell-shaped, and the normal Q–Q plot also did not show much deviation from the diagonal. The ICC for the effect of participant was .042.

There was a significant interaction between RI and practice type in all three tests: form recognition: $F_{(3,\ 633.060)} = 8.055$, $p < .001$; translation: $F_{(3,\ 655.853)} = 53.908$, $p < .001$; and translation matching: $F_{(3,\ 652.992)} = 29.186$, $p < .001$. There were no other significant interactions (all $ps > .05$). Feedback study time did not interact with any of the other independent variables in any of the tests (all $ps > .05$). Feedback study time did not have a significant main effect for the form recognition test: $F_{(1,\ 633.159)} = 1.322$, $p = .251$, $d = 0.06$. Here, the effect size is very small. A larger sample size may be needed for such a small effect to be found significant; however, its size tells us that it may not be practically interesting. Feedback study time had a significant positive main effect for the other two tests: translation, $F_{(1,\ 655.856)} = 13.606$, $p < .001$, $d = 0.21$; and translation matching, $F_{(1,\ 653.004)} = 15.241$, $p < .001$, $d = 0.23$. Here, the effect sizes are a bit larger but still very small, suggesting only a slight benefit of longer study. In evaluating effect sizes of an intervention, it is important to consider its cost. Given that such longer presentation duration takes three times more time, it may not be a good investment given the small effect sizes that may result.

To investigate the RI by practice type interaction, separate linear mixed effects analyses were run for the immediate and delayed posttests with practice type as a four-level independent variable. For consistency, all analyses were run with and without Time of Delayed Test as a covariate. Both sets of analyses showed the same pattern of results. For this reason, this covariate was excluded in order to preserve the meaningfulness of the parameter estimates in tests that could not have been affected by this covariate (immediate tests). Parameter estimates were examined with the no-practice condition and the short-spaced condition as the reference categories in two separate analyses. This allowed to compare all the levels of practice type with a minimum number of separate comparisons. For each test, the Bonferroni correction was used: $\alpha = .05/4 = .012$. Table 3 presents the parameter estimates for the effects of practice in each ISI condition. Here, the estimates are in raw percentages. The intercept represents the mean score in the no-practice condition and each slope represents the mean difference between the no-practice condition and the corresponding practice condition. The null hypothesis for the intercept is that the mean score in the no-practice condition is equal to zero. The null hypothesis for each slope is that the scores in the corresponding condition are not different from the scores in the no-practice condition. The Cohen's d effect sizes here were calculated against the baseline of no practice.

There was a significant difference between the results in the no-practice condition and each of the practice type conditions for each test type in the immediate scores. The delayed scores show a similar pattern with the exception of the translation scores in the massed condition. The slopes are positive throughout, indicating benefits of practice. However, the slopes are of different magnitudes. Thus, while the effects of retrieval practice are quite large in the two spaced conditions, the effects in the massed condition are considerably smaller. Further, the scores on the delayed translation test are not significantly different between the massed practice and the no-practice conditions at the corrected alpha level. Here, the difference between the two conditions is only roughly 5%, suggesting a negligible benefit of massed retrieval practice for L2-L1 translation ability in the long term. The Cohen's d effect sizes here need to be interpreted with the nature of the differences in mind. Because a comparison is being made between learning outcomes from six retrieval-restudy events distributed under each of the three ISI conditions and learning outcomes from no retrieval practice at all and only a single study event, we expect to see larger effects overall than when two different learning conditions that are matched with respect to variables such as time on task are compared. The effect sizes in the two spaced conditions are much larger than those in the massed condition, on all the tests. The effect of massed practice, given the nature of the comparisons, is quite small. While it is likely that with a larger sample size this effect might reach statistical significance even at the corrected alpha level, it may not be practically significant given its small size. A cost-to-benefit analysis would suggest that massed L2-L1 translation retrieval practice does not appear to be an efficient method of study, particularly for the long-term retention of knowledge, which is of primary interest in vocabulary learning.

Next, parameter estimates were examined with the short-spaced practice condition as the reference category. Here, the intercept respresents the mean score in the short-spaced practice condition and each slope represents the mean difference between this condition and the corresponding condition. The null hypothesis for the intercept here is that the mean of the scores in the short-spaced practice condition is equal to zero. The null hypothesis for each slope is that the scores in the corresponding condition are not different from the scores in the short-spaced practice condition. Table 4 presents these comparisons. On the immediate posttests, there was no significant difference between the scores in the long-spaced condition

https://doi.org/10.1017/S0142716421000370 Published online by Cambridge University Press

Table 3. Results relative to the no-practice condition

Parameter	Estimate	SE	df	t	Sig.	99.6% CI LL	99.6% CI UL	Cohen's d
Form recognition								
Immediate test								
Intercept	12.62	2.274	106.563	5.550	<.001	5.93	19.31	
Massed	19.85	2.251	317.176	8.820	<.001	13.32	26.38	1.26
Short-spaced	67.17	2.251	317.176	29.846	<.001	60.65	73.70	3.28
Long-spaced	68.36	2.244	317.075	30.465	<.001	61.85	74.86	3.92
Delayed test								
Intercept	14.89	2.416	111.159	6.164	<.001	7.79	22.00	
Massed	12.81	2.416	324.128	5.300	<.001	5.80	19.81	0.99
Short-spaced	54.19	2.416	324.128	22.426	<.001	47.18	61.19	2.70
Long-spaced	53.72	2.409	324.023	22.301	<.001	46.74	60.71	2.93
L2-L1 translation								
Immediate test								
Intercept	4.63	2.468	98.890	1.876	.064	−2.64	11.90	
Massed	13.08	2.290	330.869	5.711	<.001	6.44	19.72	1.00
Short-spaced	71.68	2.297	330.954	31.211	<.001	65.02	78.34	3.18
Long-spaced	71.30	2.297	330.954	31.043	<.001	64.64	77.95	3.11
Delayed posttest								
Intercept	2.14	2.157	139.571	0.993	.323	−4.17	8.45	
Massed	4.98	2.348	333.000	2.120	.035	−1.83	11.78	0.73
Short-spaced	35.36	2.348	333.000	15.061	<.001	28.55	42.16	1.77
Long-spaced	43.34	2.348	333.000	18.462	<.001	36.54	50.15	2.13
Form-meaning matching								
Immediate test								
Intercept	12.85	2.133	104.890	6.022	<.001	6.57	19.13	
Massed	16.67	2.051	329.629	8.125	<.001	10.72	22.61	1.02
Short-spaced	76.26	2.063	329.816	36.962	<.001	70.28	82.24	4.71
Long-spaced	75.32	2.057	329.724	36.615	<.001	69.36	81.29	4.56
Delayed test								
Intercept	4.22	2.501	99.255	1.689	.094	−3.15	11.59	
Massed	7.23	2.321	331.044	3.117	.002	0.51	13.96	0.80
Short-spaced	48.95	2.328	331.129	21.026	<.001	42.20	55.69	2.00
Long-spaced	56.88	2.328	331.129	24.435	<.001	50.13	63.63	2.86

Table 4. Results relative to the short-spaced practice condition

Parameter	Estimate	SE	df	t	Sig.	99.6% CI LL	99.6% CI UL	Cohen's d
Form recognition								
Immediate test								
Intercept	79.79	2.281	107.628	34.987	<.001	73.08	86.50	
Massed	−47.32	2.257	317.282	−20.963	<.001	−53.87	−40.78	2.48
Long-spaced	1.18	2.251	317.176	0.526	.599	−5.34	7.71	0.14
No practice	−67.17	2.251	317.176	−29.846	<.001	−73.70	−60.65	3.28
Delayed test								
Intercept	69.08	2.423	112.260	28.506	<.001	61.96	76.20	
Massed	−41.38	2.423	324.236	−17.076	<.001	−48.41	−34.36	2.11
Long-spaced	−0.46	2.416	324.128	−0.191	.848	−7.47	6.54	0.01
No practice	−54.19	2.416	324.128	−22.426	<.001	−61.19	−47.18	2.70
L2-L1 translation								
Immediate test								
Intercept	76.31	2.474	99.766	30.841	<.001	69.02	83.60	
Massed	−58.60	2.297	330.954	−25.516	<.001	−65.26	−51.95	2.81
Long-spaced	−0.39	2.303	331.042	−0.167	.867	−7.06	6.29	0.02
No practice	−71.68	2.297	330.954	−31.211	<.001	−78.34	−65.02	3.18
Delayed test								
Intercept	37.50	2.157	139.571	17.384	<.001	31.19	43.81	
Massed	−30.38	2.348	333.000	−12.941	<.001	−37.19	−23.58	1.67
Long-spaced	7.99	2.348	333.000	3.402	.001	1.18	14.79	0.45
No practice	−35.36	2.348	333.000	−15.061	<.001	−42.16	−28.55	1.77
Form-meaning matching								
Immediate test								
Intercept	89.11	2.145	106.862	41.543	<.001	82.80	95.42	
Massed	−59.60	2.063	329.816	−28.884	<.001	−65.58	−53.62	2.92
Long-spaced	−0.94	2.069	329.918	−0.453	.651	−6.94	5.06	0.04
No practice	−76.26	2.063	329.816	−36.962	<.001	−82.24	−70.28	4.71
Delayed test								
Intercept	53.17	2.507	100.134	21.208	<.001	45.78	60.56	
Massed	−41.71	2.328	331.129	−17.919	<.001	−48.46	−34.96	1.89
Long-spaced	7.94	2.335	331.218	3.399	.001	1.17	14.70	0.41
No practice	−48.95	2.328	331.129	−21.026	<.001	−55.69	−42.20	2.00

https://doi.org/10.1017/S0142716421000370 Published online by Cambridge University Press

Table 5. Training-phase response latencies and retrieval successes across the five true retrieval attempts

Practice condition	M	N	SD	Mdn	Cohen's d
Study-phase retrieval latencies					
Massed practice	6120	48	824	6078	
Short-spaced practice	14620	48	4201	14692	2.24
Long-spaced practice	16847	48	5991	14889	1.89
Study-phase retrieval success (out of five possible successful retrieval attempts)					
Massed practice	4.94	48	0.09	5.00	
Short-spaced practice	3.42	48	0.70	3.63	−2.23
Long-spaced practice	2.74	48	0.92	2.71	−2.42

and in the short-spaced condition. The difference in raw percent correct here is $<1\%$ on all measures, with the Cohen's d being also very small, suggesting a negligible effect that would likely require a very large sample size to be significant but one that would likely not be interesting in a practical sense.

On the delayed posttests, however, there was a significant advantage of long-spaced practice over short-spaced practice, indicating a significant lag effect in these scores. The massed condition produced significantly lower scores than the short-spaced condition on all the tests. In terms of the effect sizes, the difference between the long-spaced and the short-spaced condition is not large on all the tests. The difference between the massed practice and the short-spaced practice condition is quite large, on all the tests, suggesting important benefits of spacing L2-L1 retrieval practice. It is important to note that, with regard to effect sizes, spacing study produced much larger effects than did increasing feedback study time.

Moderated mediation

To answer the third and fourth research questions, moderated mediation analyses were performed with the SPSS PROCESS 3.5.2 macro (Hayes, 2018) to explore whether the dual mechanism of successful effortful retrieval during study underlies benefits of lag and whether feedback study time moderates this relationship.

The study phase produced a low percentage of errors ($M = 2.4\%$, $SD = 1.9\%$, Median $= 1.8\%$, Min $= 0.2\%$, Max $= 8.8\%$). Therefore, all participants' data were included in the analyses. Table 5 presents information on retrieval latencies and successes during study in the three ISI conditions. Effect sizes were calculated relative to the massed condition. The shortest latencies were observed in the massed condition. The short-spaced condition produced latencies that were twice as long as those in the massed condition and the long-spaced condition produced only slightly longer latencies than the short-spaced condition. It is important to note that because the values here include both successful and unsuccessful retrieval attempts mixed together, response latencies in the long-spaced condition were likely affected by the fact that at such long ISIs words may often not have been recognized as ones studied previously, in which case participants produced "I don't know" responses

https://doi.org/10.1017/S0142716421000370 Published online by Cambridge University Press

Table 6. Training-phase response latencies and retrieval successes in the two study time conditions

Practice condition	M	N	SD	Mdn	Cohen's d
Shorter feedback study time					
Study-phase retrieval latencies					
Massed practice	6175	48	940	5983	
Short-spaced practice	14735	48	4639	14707	
Long-spaced practice	17200	48	6116	15854	
Study-phase retrieval success (out of five possible successful retrieval attempts)					
Massed practice	4.92	48	0.12	5.00	
Short-spaced practice	3.31	48	0.76	3.42	
Long-spaced practice	2.54	48	0.97	2.42	
Longer feedback study time					
Study-phase retrieval latencies					
Massed practice	6065	48	832	5899	0.17
Short-spaced practice	14505	48	4466	13618	0.07
Long-spaced practice	16494	48	6540	14144	0.17
Study-phase retrieval success (out of five possible successful retrieval attempts)					
Massed practice	4.95	48	0.10	5.00	0.29
Short-spaced practice	3.54	48	0.78	3.67	0.36
Long-spaced practice	2.94	48	1.02	3.00	0.53

without engaging in a search of their memory, which resulted in faster responses to these words. Retrieval in the massed condition was almost always successful. Retrieval success decreased with spacing: in the short-spaced condition there were fewer successful retrieval events and in the long-spaced condition these were even fewer.

Table 6 presents these results separately for the long and short feedback study time conditions. Effect sizes for ISI were calculated relative to the massed practice condition. The effects of ISI seem a bit more pronounced in the shorter study time condition than in the longer study time condition. This makes intuitive sense, as longer study time should increase retrieval success and decrease retrieval effort, particularly with spaced practice. Words that were presented for study for 9 s received slightly less overall retrieval effort and slightly more retrieval success. Separate analyses further showed that the overall pattern of study phase response latencies and accuracy was similar between the two genders.

Data reduction was performed to reduce the six sets of scores to fewer dependent variables for the moderated mediation analyses. Based on correlations, theoretical reasons, and principal component analyses, three dependent variables emerged. These combined together (1) the immediate and delayed form recognition tests, (2) the two immediate meaning tests, and (3) the two delayed meaning tests.

Table 7. Correlation coefficients and principal component analysis results

Pearson correlation coefficients				Principal component analysis
Form recognition tests		Immediate form recognition	Delayed form recognition	Component 1 (83.5 % variance explained)
	Immediate form recognition	1		.914
	Delayed form recognition	.671***	1	.914
Immediate meaning tests		Immediate translation test	Immediate form-meaning mapping test	Component 1 (95.0 % variance explained)
	Immediate translation test	1		.975
	Immediate form-meaning mapping test	.900***	1	.975
Delayed meaning tests		Delayed translation test	Delayed form-meaning m apping test	Component 1 (94.5 % variance explained)
	Delayed translation test	1		.972
	Delayed form-meaning mapping test	.890***	1	.972

***$p < .001$.

The three resulting sets of scores will be named, respectively, the form recognition tests, the immediate meaning tests, and the delayed meaning tests. Table 7 presents the bivariate correlations between each member of a pair as well as loadings of each pair of tests on their corresponding extracted component. Each analysis shows high loadings, suggesting that the corresponding test pair likely measures the same underlying construct.

Moderated parallel mediation analyses

Because multiple models were run on the same or related data, the alpha level was corrected. Further, robust tests were used to ensure against any violations of normality. Bootstrapped 99% confidence intervals (99% BCIs) were requested with 10,000 bootstrap samples. An initial model investigated whether the two mechanisms of retrieval effort and success underlie lag effects and whether feedback study time affects the operation of these two mechanisms. The moderated parallel mediation analysis included effort and success during training as mediators and feedback study time as a moderator of the relationship between ISI and the two mediators and ISI and learning outcomes (Model 8). Time of Delayed Test was included as a

https://doi.org/10.1017/S0142716421000370 Published online by Cambridge University Press

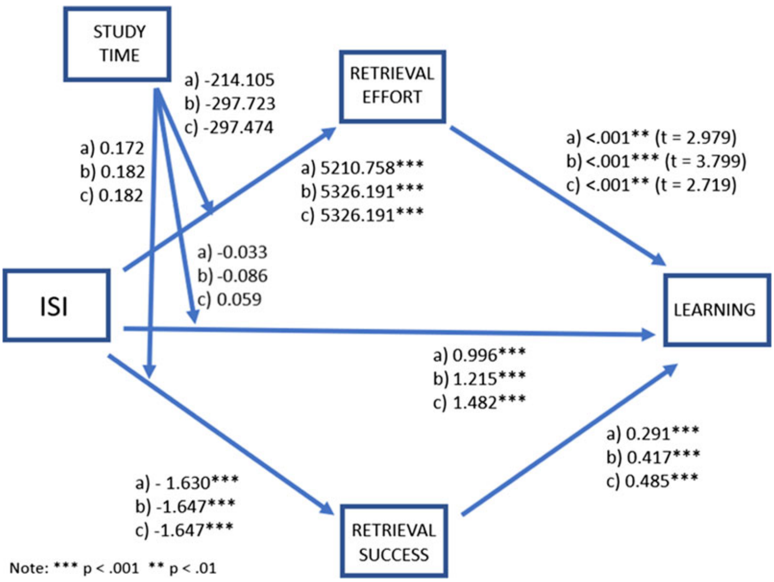

Figure 5. The conceptual structure of the moderated parallel mediation analysis. The form recognition, L2-L1 translation, and translation matching tests are denoted as a, b, and c, respectively.

covariate. Because the homogeneity of slopes assumption was violated in at least one test, the interaction between the covariate and the corresponding variable (ISI) was included as the covariate in all tests for consistency. Figure 5 presents the conceptual structure of this analysis with the obtained coefficients.

ISI had a significant positive effect on learning in all three tests. It further had a significant positive effect on retrieval effort and a significant negative effect on retrieval success in all three tests. Both retrieval effort and retrieval success significantly positively affected learning in all three tests. Feedback study time did not significantly moderate the relationships between ISI and retrieval effort, ISI and retrieval success, or ISI and learning outcomes in any of the tests. While the moderating effects of this variable are in the predicted direction, these effects are very small relative to the main effects of ISI on retrieval effort and success during training. Thus, the effects of ISI on retrieval effort and success were not moderated by the level of study time to a substantial degree.

For all three sets of scores, there was significant mediation by retrieval success as a negative effect, across the two levels of feedback study time: the form recognition tests: $\beta = -.4236$, bootstrapped $SE = .1070$, 99% BCI [−.7125, −.1578] for short presentation duration and $\beta = -.3736$, bootstrapped $SE = .0968$, 99% BCI [−.6329, −.1352] for long presentation duration; the immediate meaning tests: $\beta = -.6111$, bootstrapped $SE = .0971$, 99% BCI [−.8739, −.3654] for short presentation duration and $\beta = -.5352$, bootstrapped $SE = .0903$, 99% BCI [−.7833, −.3174] for long presentation duration; and the delayed meaning tests: $\beta = -.7107$, bootstrapped $SE = .1094$, 99% BCI [−1.0120, −.4357] for short presentation duration and $\beta = -.6224$, bootstrapped $SE = .1010$, 99% BCI [−.9055, −.3749] for long presentation duration. This suggests that, despite the fact that a

https://doi.org/10.1017/S0142716421000370 Published online by Cambridge University Press

nonmonotonic function of lag was not observed in the present results, a negative effect of longer ISI was still present and operated through a lower rate of retrieval success during study.

For all three sets of scores, there was significant mediation by retrieval effort as a positive effect, across the two levels of feedback study time: the form recognition tests: $\beta = .1486$, bootstrapped $SE = .0535$, 99% BCI [.0303, .3142] for short presentation duration and $\beta = .1423$, bootstrapped $SE = .0494$, 99% BCI [.0285, .2902] for long presentation duration; the immediate meaning tests: $\beta = .1668$, bootstrapped $SE = .0481$, 99% BCI [.0577, .3098] for short presentation duration and $\beta = .1569$, bootstrapped $SE = .0430$, 99% BCI [.0545, .2810] for long presentation duration; and the delayed meaning tests, $\beta = .1253$, bootstrapped $SE = .0484$, 99% BCI [.0019, .2572] for short presentation duration and $\beta = 1179$, bootstrapped $SE = .0491$, 99% BCI [.0018, .2618] for long presentation duration. However, there was no significant moderated mediation in any of the three sets of scores: the form recognition tests: Index of Moderated Mediation (success) = .0499, bootstrapped $SE = .0326$, 99% BCI [−.0311, .1458] and Index of Moderated Mediation (effort) = −.0064, bootstrapped $SE = .0222$, 99% BCI [−.0778, .1519]; the immediate meaning tests: Index of Moderated Mediation (success) = .0759, bootstrapped $SE = .0425$, 99% BCI [−.0323, .1926] and Index of Moderated Mediation (effort) = −.0099, bootstrapped $SE = .0232$, 99% BCI [−.0793, .0481]; and the delayed meaning tests: Index of Moderated Mediation (success) = .0883, bootstrapped $SE = .0491$, 99% BCI [−.0401, .2242] and Index of Moderated Mediation (effort) = −.0074, bootstrapped $SE = .0174$, 99% BCI [−.0536, .0506].

Note that both retrieval effort and retrieval success were modeled in this analysis as main effects. However, a dual mechanism of effortful successful retrieval implies an interaction, where the effect of one may depend on the level of the other. The question whether the positive effects of retrieval effort are conditional on retrieval success will be tested in the following moderated mediation analysis.

Mediation by retrieval effort moderated by retrieval success (a moderated mediation analysis)

Retrieval effort was chosen as the mediator of the effects of spacing on learning. Retrieval success was chosen as a moderator of this mediation. The reason for this choice was theoretical. Because retrieval effort is known to promote word learning (e.g., Pyc & Rawson, 2009), it is an interesting question whether the benefits of increased effort that results from longer ISIs in retrieval practice are conditional on retrieval success. It is further interesting to know whether this holds in the presence of feedback that follows each retrieval attempt. Provision of feedback after each retrieval attempt is a more usual situation for second-language vocabulary learning. The moderated parallel mediation analysis showed that despite the fact that a nonmonotonic function was not observed in the learning outcomes, retrieval failure during study that resulted from spacing retrieval attempts more widely still had a negative effect on learning. It is an important question whether retrieval success moderates beneficial effects of retrieval effort on learning and may thus constitute a limitation for how widely we may space retrieval practice even in the presence of feedback.

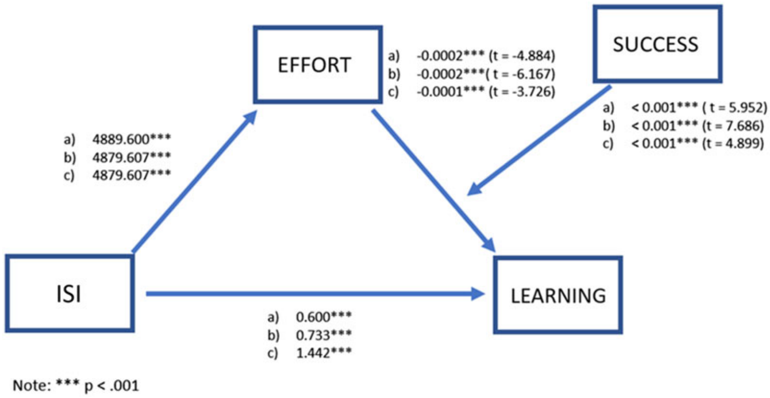

Figure 6. The conceptual structure of the moderated mediation analysis. The form recognition, L2-L1 translation, and translation matching tests are denoted as a, b, and c, respectively.

Because feedback study time was found to only have a small nonsignificant moderating effect on the relationship between ISI and retrieval effort and success during study, participants' scores were collapsed across the levels of this variable for this analysis. Figure 6 presents the conceptual structure of the moderated mediation analysis (Model 14) with the obtained coefficients. The coefficients show a similar pattern for all three sets of vocabulary scores. There is a positive effect of ISI on retrieval effort and also on the learning outcomes. The effect of effort is now actually negative in each of the three sets of vocabulary scores. However, retrieval success significantly positively moderates this relationship.

The tests of the indirect effects showed significant moderated mediation for all learning measures: the form recognition tests: Index of Moderated Mediation = .3239, bootstrapped *SE* = .0846, 99% BCI [.1588, .5812]; the immediate meaning tests: Index of Moderated Mediation = .3494, bootstrapped *SE* = .0721, 99% BCI [.2010, .5734]; and the delayed meaning tests: Index of Moderated Mediation = .2443, bootstrapped *SE* = .0771, 99% BCI [.0930, .4830]. To investigate more in depth the moderated mediation process, the effect of the mediator was tested at different levels of the moderator variable, in this case, using the 16th, 50th, and 84th percentiles. Table 8 presents the effect of retrieval effort on vocabulary scores at the three levels of retrieval success represented by the three percentiles. This table shows a similar pattern across the three sets of scores. It shows that the effect of retrieval effort becomes positive and grows in magnitude as retrieval success rate increases. This indicates that the beneficial effects of effort were contingent on a higher rate of retrieval success in this moderated mediation analysis.

Discussion

The present research examined the contribution of the dual mechanism of successful effortful retrieval during study to the effects of spacing L2-L1 retrieval practice on learning novel L2 vocabulary within a declarative knowledge acquisition task. It further investigated the effects of feedback study time, per encounter and in total,

https://doi.org/10.1017/S0142716421000370 Published online by Cambridge University Press

Table 8. Effect of retrieval effort at three levels of retrieval success

Retrieval success rate	Effect of retrieval effort	BootSE	BootLLCI	BootULCI
Form recognition tests				
2.33	−0.21	0.1042	−0.53	0.01
3.75	0.25	0.0905	0.07	0.52
5.00	0.66	0.1724	0.32	1.19
Immediate meaning tests				
2.33	−0.20	0.0874	−0.46	−0.01
3.75	0.30	0.0643	−0.16	0.49
5.00	0.73	0.1356	0.45	1.15
Delayed meaning tests				
2.33	−0.10	0.0974	−0.40	0.14
3.75	0.25	0.0758	0.09	0.48
5.00	0.55	0.1494	0.26	1.02

on learning outcomes and on the operation of the two study-phase mechanisms under investigation.

The first research question asked whether spacing practice more widely affects learning outcomes, as measured by immediate and delayed form recognition and translation posttests. The results showed a spacing effect of considerable size across the posttest types and RIs. Importantly, the difference between the massed practice condition and the no-practice control condition was very small, particularly in terms of the long-term gains, where, on the most challenging L2-L1 translation test, scores in the massed practice condition were not significantly different from those in the no-practice condition. This suggests that, despite the fact that retrieval practice is believed to be beneficial, massed retrieval practice may not be an effective learning tool if we are targeting longer-term retention of knowledge, which is usually more relevant for L2 vocabulary learning. The present findings are in line with proposals that retrieval from short-term memory may not involve processes that make retrieval beneficial for learning (Glover, 1989). An anonymous reviewer has pointed out that, because the immediate tests were conducted 30 min after the study phase in this experiment, the scores on these tests may not be considered immediate scores in the strictest sense. Thus, in the present study, the benefits of massed practice, which are usually most pronounced when knowledge is measured immediately after study, may be underestimated in terms of their immediate effects due to the operationalization of immediate learning as scores on a test that followed the study phase after a bit of a delay.

Retrieval practice was distributed under three levels of lag. The results showed a significant lag effect (advantage of long-spaced practice over short-spaced practice) on the delayed meaning posttests but not on the immediate meaning posttests. However, no lag effect (but only a spacing effect) was observed on both form recognition posttests, where the scores in the short- and long-spaced conditions

https://doi.org/10.1017/S0142716421000370 Published online by Cambridge University Press

were similar. This pattern of results is in line with previous findings of more pronounced beneficial effects of lag the more challenging the task (Maddox, 2016) and with findings that effects of spacing study more widely become more pronounced when knowledge is tested after a longer period of time (Delaney, Verkoeijen, & Spirgel, 2010; Nakata, 2015; Nakata & Webb, 2016). Thus, the results suggest that the temporal distribution of retrieval practice may be crucial: massed practice may be not much better than no practice at all but only a single study event, in the long term, and longer intervals between repetitions may produce more robust knowledge that is forgotten more slowly.

The second research question asked whether feedback study time (3 vs. 9 s) affects learning outcomes. Prior research has shown that learners are not effective at pacing their own study (Rundus, 1971), often devoting more study time to items that they believe to be more difficult, such as to spaced rather than massed repetitions, when this impression may not always be accurate. It was an interesting question whether longer study time that is imposed externally can counteract negative consequences of massed study. The present results showed that longer study time has a significant, though quite small, positive effect for knowledge of meaning though not of form. The small size of the effect is different from the considerable learning benefits of more attentional processing found in prior SLA research (e.g., Godfroid et al., 2018; Koval, 2019). An important difference may be that in such prior research learners were free to self-pace their study. The present findings suggest that when longer study time is imposed externally, it may not have benefits of the same magnitude as when a learner chooses to devote longer study time to a target word. This, in turn, suggests that the processes that underlie self-regulated and other-imposed longer study time are qualitatively different. Recall that the time participants were given for studying a word in the longer study time condition was three times longer than that in the shorter study time condition. However, the benefit that came from tripling the study time was quite small.

The third and fourth research questions asked whether the dual mechanism of successful effortful retrieval during study underlies benefits of lag and whether feedback study time moderates this relationship, respectively. The results of the moderated mediation analyses showed that increasing feedback study time from 3 to 9 s had a small, nonsignificant effect on the operation of the two cognitive mechanisms under investigation. They further showed that, despite the fact that a nonmonotonic function of lag was not observed in the present learning outcomes, a negative effect of increasing ISI was still present and operated through a lower rate of retrieval success during study. Further, on all learning measures, benefits of retrieval effort were conditional on retrieval success, in line with the predictions of the reminding account.

Despite the fact that in the present study each retrieval attempt was followed by feedback, failed retrieval attempts did not benefit from more effort. This is surprising because, with failed retrieval attempts, one would expect a more effortful search of one's memory to result in higher quality processing of subsequently presented feedback, which should, in turn, benefit learning (Izawa, 1970; Kornell et al., 2009). Further, despite the study of feedback, spacing repeated retrieval-restudy events more widely had a negative effect on learning by negatively affecting retrieval success during study. This finding disconfirms proposals that training-phase

https://doi.org/10.1017/S0142716421000370 Published online by Cambridge University Press

retrieval failures that result from spacing practice more widely are beneficial for learning (Bahrick & Hall, 2005; Pashler et al., 2003). In the latter research, the effects of retrieval failure that resulted from longer spacing were not directly tested but only inferred based on the finding that the longest-spaced condition produced most frequent retrieval failures during study but also superior learning outcomes. On the surface, the same pattern seems to hold in the present results: the long-spaced condition produced the lowest study-phase performance success and the highest posttest scores in the long term. However, the moderated mediation analyses allowed to disentangle these complex relationships more effectively and to detect negative effects of longer spacing. The present results suggest that a balance must be struck between study-phase performance effort and success: words that are retrieved successfully though with difficulty during the study phase are retained better than those that are not retrieved or are retrieved with minimal effort.

Pedagogical implications

The findings of the present research have important implications for second-language vocabulary teaching and learning. The findings indicate, first of all, that despite the fact that retrieval practice is believed to be beneficial, how closely together or widely apart retrieval events occur has very important consequences for L2 vocabulary learning outcomes. If retrieval events occur consecutively or in very close succession, such practice may have little to no benefits for longer-term retention. Despite the fact that the control condition did not involve any true retrieval attempts and only involved one study event, whereas massed practice involved five true (and predominantly successful) retrieval events and six times longer total study of a translation pair, the difference in learning outcomes between these two conditions was very small, particularly in the long term. This finding further suggests that increasing the number of retrieval-(re)study events that occur consecutively or closely together (even if this is increased from one to six) has very little benefit and may not be a good way to use study time. Learners are known to sometimes engage in such self-drilling, whereby they repeat a given word with its translation for a considerable length of time, believing that the longer they rehearse it the better it will be remembered; or test themselves on an item that was very recently seen and while retrieval is still very easy because the information still resides in working memory. The present findings suggest that there may be little to no benefit of such drilling or massed retrieval practice over a single short study event. Learners should schedule self-testing repetitions such that retrieval of the studied material is attempted only once they feel that some, though not complete, forgetting has occurred. For example, if a learner is studying 20 words with their translations, they may wish to go through the entire list before revisiting any given item rather than devoting a number of consecutive retrieval-restudy events to the same item before moving on to the next item. To use time more efficiently, learners may also wish to cut study of the same item short as soon as they feel that it has been sufficiently encoded in memory, if it is to be revisited repeatedly.

Longer intervals between within-study-session retrieval attempts can be used to enhance learning from retrieval practice and slow forgetting of learned material. The higher retrieval effort that results from longer intervals between repetitions underlies these benefits of more widely spaced retrieval practice. However, the benefit of increased retrieval effort is conditional on retrieval success. When selecting a retrieval practice schedule, such as for word learning software or materials design, we need to take into account the probability of retrieval effort and success given our specific circumstances and learner variables. Many different variables may affect retrieval effort and the probability of retrieval success during the training phase. These may be the difficulty of the studied material, the age group and memory ability of our learners, and/or the complexity and interference potential of the intervening material or activity. Thus, we may want to use shorter ISIs with more difficult or complex tasks, for example, in order to ensure a higher rate of retrieval success during study.

Increasing the time, per encounter and in total, that a learner is given to study an L2 word presented with its meaning, such as longer presentation rate in PAL software, has a small beneficial effect on memory and also slightly increases the chances of successful retrieval during study. Increasing study time does not, however, counteract the negative effects of using massed instead of spaced practice, even if such practice involves retrieval. Previous research has shown that more attentional processing of target words leads to more learning (Godfroid et al., 2018; Godfroid, et al., 2013) and may be the reason spacing repeated study results in superior learning outcomes (Koval, 2019; Rundus, 1971). The present results suggest that the large benefits of longer study time may be limited to situations where learners choose to allocate more study time to a target item based on learner-internal reasons and may not be observed when longer study time is externally imposed on the learner. This suggests that our efforts should be aimed at getting learners to choose to allocate more attention or study time to target forms, such as, for example, by using spacing (Koval, 2019), rather than imposing longer study time externally. At least for receptive knowledge development, computer programs that present immediate feedback after each retrieval attempt need not make feedback presentation longer than is reasonably enough for successful encoding of the information (without additional time to simply rehearse), as doing so appears not to have large benefits and may, therefore, not represent efficient use of time.

Finally, the results suggest that if there is a chance that a learner may be able to retrieve a given target piece of information from memory, they should be allowed to take the time they need to do so rather than being presented with the information before the retrieval process is complete. It is often tempting, in the interest of time, to present information that a learner might otherwise take a longer time to retrieve on their own. However, if we rush to present the target information before a learner completes a potentially successful retrieval attempt, this may constitute a less powerful learning event than if the information were fully retrieved from memory.

Limitations and suggestions for future research

An important limitation of the within-participant design adopted in this study is the fact that the participants completed the same posttests twice. The retrieval processes that occurred during the immediate posttests may have had an effect on the performance on the delayed posttests. However, all the conditions had an equal chance to benefit from such additional practice. Another limitation of the study is that the number of words in each sublist is on the small side, which was done in order not to overwhelm the participants with the number of words they needed to learn.

The present study investigated the contribution of the dual mechanism of successful effortful retrieval to the lag effect in L2 vocabulary learning within a declarative knowledge acquisition task. Retrieval was operationalized as overt retrieval of L1 translations for target L2 words in a paired-associate learning format. It is important to note, however, that overt L2-L1 translation retrieval is only one type of retrieval practice and only one type of retrieval. Overt retrieval is pedagogically interesting primarily because it can be observed. It is an important question whether we need to schedule repeated retrieval events such that they are effortful but still mostly successful, a question that leads to very straightforward pedagogical recommendations. Future research needs to supplement the present results with an investigation of L1-L2 retrieval practice. Such an investigation is also likely to result in pedagogical recommendations that can be applied with relative ease. L1-L2 translation is a more challenging task, which is likely to mean dramatically less retrieval success during study at longer ISIs. Because retrieval failure was shown in the present experiment to have a negative effect on learning and also to interfere with beneficial effects of retrieval effort, an investigation of L1-L2 retrieval practice may be more likely to capture a nonmonotonic function of lag in learning outcomes, whereby longer-spaced practice may actually produce inferior results to shorter-spaced practice, which was not observed in the present experiment. Thus, for example, using L1-L2 translation practice, Nakata (2015) showed a beneficial effect of spacing but not of lag. The benefits of lag observed in the present study may, at least in part, be due to the fact that L2-L1 translation is an easier task.

In learning situations that do not involve overt retrieval, benefits of spacing may still depend on a covert retrieval process. Future studies need to explore the contribution of covert retrieval to spacing and lag effects in such learning tasks as well. Such covert retrieval can be observed through tests of simple recognition or through indirect memory tests such as facilitation, or speed-up, in task performance. For example, in Koval (2019), I used eye-tracking to examine facilitation in reading times for L2 words encountered multiple times within sentence contexts during study as an indication of a covert retrieval process. However, here I did not intentionally attempt to manipulate retrieval success during study, but only explored this post hoc. Future studies should aim to induce retrieval failure at longer ISIs during performance of L2 learning tasks that do not involve overt retrieval in order to investigate the mediating effects of covert retrieval success.

In the present study, retrieval practice and intentional learning within a declarative knowledge acquisition task may have resulted in stronger memory traces established at each repetition, which may have led to better study-phase performance. This may be one reason why a nonmonotonic function of lag was not observed in the present results. A nonmonotonic function may be easier to capture in a task that may not establish very strong memory traces, such as, for example, incidental learning of vocabulary from reading comprehension activities (Verkoeijen, et al., 2005). Future research will also need to examine longer ISIs to capture this function more effectively. Thus, for example, despite the fact that participants continued retrieval attempts until correct performance during the training phase in Cepeda et al. (2009, Exp 1), the results showed a nonmonotonic function in the learning

https://doi.org/10.1017/S0142716421000370 Published online by Cambridge University Press

outcomes 10 days after study that was distributed under six different levels of lag, the longest lag being 14 days' ISI. An anonymous reviewer has pointed out that many of the SLA studies that have failed to observe benefits of longer ISIs have involved learning over multiple sessions. This may well be due to the fact that more dramatic retrieval failures are produced with such longer ISIs. Another difference between within- and between-session study is that the latter involves sleep-associated consolidation. Future studies will need to compare directly learning from within- and between-session repetitions and investigate how this affects the operation of the underlying cognitive mechanisms.

The present research investigated the effects of externally predetermined feedback study time. Feedback study time is only one potentially relevant variable that may affect the operation of the underlying mechanisms of retrieval effort and success. Other relevant variables are numerous. The issue of what variables will affect the probability of performance success in different learning situations is still unresolved and warrants further investigation. These effects need to be investigated for a fuller picture of the conditions under which various amounts of spacing may be beneficial or detrimental for L2 learning. In the present study, participants studied novel L2 words that represented simple and generic concepts, in a completely novel language, from six repeated L2-L1 retrieval attempts followed by feedback, within one study session. Future research needs to examine other tasks and learning contexts and other learning targets, as well as other learner proficiencies. It will be important also to test the effects of different numbers of repetitions: it may be that fewer repetitions are needed with spaced practice (Maddox & Balota, 2015); however, this may, in turn, depend on other relevant variables and their effects on the cognitive mechanisms that underlie the effects of spacing study of L2 material.

Acknowledgments. This research was supported in part by the Second Language Studies doctoral program at Michigan State University, USA. I would like to thank my dissertation committee members Drs. Charlene Polio, Patti Spinner, Susan Gass, Aline Godfroid, and Sandra Deshors as well as my three anonymous reviewers and Associate Editor Dr. Miquel Simonet for their valuable comments and suggestions. Any errors that remain are my own.

Dedication. In loving memory of my father, Grigory Ivanovich Koval

References

Arnold, K. M., & McDermott, K. B. (2013). Test-potentiated learning: Distinguishing between direct and indirect effects of tests. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **39**(3), 940–945.

Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1995). The CELEX Lexical Database. Release 2 [CD-ROM]. Linguistic Data Consortium, University of Pennsylvania, Philadelphia.

Bahrick, H. P., Bahrick, L. E., Bahrick, A. S., & Bahrick, P. E. (1993). Maintenance of foreign language vocabulary and the spacing effect. *Psychological Science*, **4**, 316–321.

Bahrick, H. P., & Hall, L. K. (2005). The importance of retrieval failures to long-term retention: A metacognitive explanation of the spacing effect. *Journal of Memory & Language*, **52**(4), 566–577.

Barcroft, J. (2007). Effects of opportunities for word retrieval during second language vocabulary learning. *Language Learning*, **57**, 35–56.

Benjamin, A. S., & Ross, B. H. (2010). The causes and consequences of reminding. In A. S. Benjamin (Ed.), *Successful remembering and successful forgetting: A Festschrift in honor of Robert A. Bjork* (pp. 71–87). New York: Psychology Press.

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

Benjamin, A.S., Bjork, R.A., & Schwartz, B.L. (1998). The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General*, **127**, 55–68.

Bird, S. (2010). Effects of distributed practice on the acquisition of second language English syntax. *Applied Psycholinguistics*, **31**, 635–650.

Bjork, R. A. (1975). Retrieval as a memory modifier. In R. Solso (Ed.), *Information processing and cognition: The Loyola Symposium* (pp. 123–144). Erlbaum.

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.

Bjork, R. A. (1999). Assessing our own competence: Heuristics and illusions. In D. Gopher & A. Koriat (Eds.), *Attention and performance* (pp. 435–459). The MIT Press.

Bjork, R. A., & Allen, T. W. (1970). The spacing effect: Consolidation or differential encoding? *Journal of Verbal Learning & Verbal Behavior*, **9**(5), 567–572.

Bloom, K.C., & Shuell, T.J. (1981). Effects of massed and distributed practice on the learning and retention of second-language vocabulary. *The Journal of Educational Research*, **74**, 245–248.

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, **46**(3), 904–911.

Bui, D. C., Maddox, G. B., & Balota, D. A. (2013). The roles of working memory and intervening task difficulty in determining the benefits of repetition. *Psychonomic Bulletin & Review*, **20**(2), 341–347.

Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition*, **20**(6), 633–642.

Cepeda, N. J., Coburn, N., Rohrer, D., Wixted, J. T., Mozer, M. C., & Pashler, H. (2009). Optimizing distributed practice: Theoretical analysis and practical implications. *Experimental Psychology*, **56**(4), 236–246.

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.

Cepeda, N. J., Vul, E., Rohrer, D., Wixted, J. T., & Pashler, H. (2008). Spacing effects in learning: A temporal ridgeline of optimal retention. *Psychological Science*, **19**(11), 1095–1102.

Challis, B. H. (1993). Spacing effects on cued-memory tests depend on level of processing. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 389–396.

Chen, O., Paas, F., & Sweller, J. (2021). Spacing and interleaving effects require distinct theoretical bases: a systematic review testing the cognitive load and discriminative-contrast hypotheses. *Educational Psychology Review*, 1–24. https://doi.org/10.1007/s10648-021-09613-w

Collins, L., Halter, R. H., Lightbown, P. M., & Spada, N. (1999). Time and the distribution of time in L2 instruction. *TESOL Quarterly*, **33**(4), 655–680.

Crowder, R.G. (1976). *Principles of learning and memory*. Hillsdale, NJ: Erlbaum.

Cull, W.L., Shaughnessy, J.J., & Zechmeister, E.B. (1996). Expanding understanding of the expanding-pattern-of-retrieval mnemonic: Toward confidence in applicability. *Journal of Experimental Psychology: Applied*, **2**, 365–378.

D'Agostino, P. R., & DeRemer, P. (1973). Repetition effects as a function of rehearsal and encoding variability. *Journal of Verbal Learning & Verbal Behavior*, **12**(1), 108–113.

Davis, C. J. (2005). N-Watch: A program for deriving neighborhood size and other psycholinguistic statistics. *Behavior Research Methods*, **37**(1), 65–70.

Delaney, P. F., Verkoeijen, P. P., & Spirgel, A. (2010). Spacing and testing effects: A deeply critical, lengthy, and at times discursive review of the literature. *Psychology of Learning and Motivation*, **53**, 63–147.

Dellarosa, D., & Bourne, L. E. (1985). Surface form and the spacing effect. *Memory & Cognition*, **13**, 529–537.

Dempster, F.N. (1988). The spacing effect: A case study in the failure to apply the results of psychological research. *American Psychologist*, **43**, 627–634.

Dempster, F. N. (1989). Spacing effects and their implications for theory and practice. *Educational Psychology Review*, **1**, 309–330.

Donovan, J. J., & Radosevich, D. J. (1999). A meta-analytic review of the distribution of practice effect: Now you see it, now you don't. *Journal of Applied Psychology*, **84**, 795–805.

Elgort, I., & Warren, P. (2014). L2 Vocabulary learning from reading: Explicit and tacit lexical knowledge and the role of learner and item variables. *Language Learning*, **64**, 365–414.

Estes, W. K. (1955). Statistical theory of distributional phenomena in learning. *Psychological Review*, **62**, 369–377.

Forster, K. I., & Forster, J. (2003). DMDX: A windows display program with millisecond accuracy. *Behavioral Research Methods, Instruments & Computers*, **35**, 116–124.

Gerbier, E., & Toppino, T. C. (2015). The effect of distributed practice: Neuroscience, cognition, and education. *Trends in Neuroscience & Education*, **4**(3), 49–59.

Glenberg, A. M. (1979). Component-levels theory of the effects of spacing of repetitions on recall and recognition. *Memory & Cognition*, **7**, 95–112.

Glenberg, A. M., & Smith, S. M. (1981). Spacing repetitions and solving problems are not the same. *Journal of Verbal Learning & Verbal Behavior*, **20**(1), 110–119.

Glover, J. A. (1989). The" testing" phenomenon: Not gone but nearly forgotten. *Journal of Educational Psychology*, **81**(3), 392–399.

Godfroid, A., Ahn, J., Choi, I., Ballard, L., Cui, Y., Johnston, S., Lee, S., Sarkar, A. & Yoon, H. J. (2018). Incidental vocabulary learning in a natural reading context: An eye-tracking study. *Bilingualism: Language & Cognition*, **21**(3), 563–584.

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

Greene, R. L. (1989). Spacing effects in memory: Evidence for a two-process account. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **15**(3), 371–377.

Hayes, A. F. (2018). *Introduction to mediation, moderation, and conditional process analysis* (2nd ed.). Guilford Press.

Hays, M. J., Kornell, N., & Bjork, R. A. (2013). When and why a failed test potentiates the effectiveness of subsequent study. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **39**(1), 290–296.

Hintzman, D. L. (1974). Theoretical implications of the spacing effect. In R. L. Solso (Ed.), *Theories in cognitive psychology: The Loyola Symposium* (pp. 77–99). Erlbaum.

Hintzman, D. L. (2004). Judgment of frequency versus recognition confidence: Repetition and recursive reminding. *Memory & Cognition*, **32**(2), 336–350.

Hintzman, D. L. (2010). How does repetition affect memory? Evidence from judgments of recency. *Memory & Cognition*, **38**(1), 102–115.

Hogan, R.M., & Kintsch,W. (1971). Differential effects of study and test trials on long-term recognition and recall. *Journal of Verbal Learning & Verbal Behavior*, **10**, 562–567.

IBM Corp. Released 2017. IBM SPSS Statistics for Windows, Version 25.0. Armonk, NY: IBM Corp.

Izawa, C. (1970). Optimal potentiating effects and forgetting-prevention effects of tests in paired-associate learning. *Journal of Experimental Psychology*, **83**, 340–344.

Jacoby, L. L. (1978). On interpreting the effects of repetition: Solving a problem versus remembering a solution. *Journal of Verbal Learning & Verbal Behavior*, **17**(6), 649–667.

Jacoby, L. L., Bjork, R. A., & Kelley, C. M. (1994). Illusions of comprehension and competence. In D. Druckman & R. A. Bjork (Eds.), *Learning, remembering, believing: Enhancing team and individual performance* (pp. 57–80). National Academy Press.

Kang, S. (2016). Spaced repetition promotes efficient and effective learning: Policy implications for instruction. *Policy Insights from the Behavioral & Brain Sciences*, **3**(1), 12–19.

Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science*, **319**, 966–968.

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

Kiliç, A., Hoyer, W. J., & Howard, M. W. (2013). Effects of spacing of item repetitions in continuous recognition memory: Does item retrieval difficulty promote item retention in older adults?. *Experimental Aging Research*, **39**(3), 322–341.

Kolers, P.A., & Roediger, H.L., III. (1984). Procedures of mind. *Journal of Verbal Learning & Verbal Behavior*, **23**, 425–449.

Kornell, N., & Bjork, R. A. (2007). The promise and perils of self-regulated study. *Psychonomic Bulletin & Review*, **14**(2), 219–224.

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

Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **35**(4), 989–998.

Koval, N.G. (2019). Testing the deficient processing account of the spacing effect in L2 vocabulary learning: Evidence from eye-tracking. *Applied Psycholinguistics*, **40**, 1103–1139.

Küpper-Tetzel, C. E., & Erdfelder, E. (2012). Encoding, maintenance, and retrieval processes in the lag effect: A multinomial processing tree analysis. *Memory*, **20**, 37–47.

Landauer, T. K. (1969). Reinforcement as consolidation. *Psychological Review*, **76**(1), 82–96.

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

Lee, I. H., Maechtle, C., & Hu, C. F. (2021). Enhancing vocabulary retention in low-achieving EFL students: Massed or spaced? *English Teaching & Learning*, 1–16. https://doi.org/10.1007/s42321-020-00074-y

Maddox, G. B. (2016). Understanding the underlying mechanism of the spacing effect in verbal learning: A case for encoding variability and study-phase retrieval. *Journal of Cognitive Psychology*, **28**(6), 684–706.

Maddox, G. B., & Balota, D. A. (2015). Retrieval practice and spacing effects in young and older adults: An examination of the benefits of desirable difficulty. *Memory & Cognition*, **43**(5), 760–774.

Maddox, G. B., Pyc, M. A., Kauffman, Z. S., Gatewood, J. D., & Schonhoff, A. M. (2018). Examining the contributions of desirable difficulty and reminding to the spacing effect. *Memory & Cognition*, **46**(8), 1376–1388.

Madigan, S. A. (1969). Intraserial repetition and coding processes in free recall. *Journal of Verbal Learning & Verbal Behavior*, **8**(6), 828–835.

McDaniel, M. A., & Fisher, R. P. (1991). Tests and test feedback as learning sources. *Contemporary Educational Psychology*, **16**, 192–201.

McDaniel, M.A., Friedman, A., & Bourne, L.E. (1978). Remembering the levels of information in words. *Memory & Cognition*, **6**, 156–164.

Melton, A.W. (1970). The situation with respect to the spacing of repetitions and memory. *Journal of Verbal Learning & Verbal Behavior*, **9**, 596–606.

Miles, S., & Kwon, C. J. (2008). Benefits of using CALL vocabulary programs to provide systematic word recycling. *English Teaching*, **63**(1), 199–216.

Miles, S. W. (2014). Spaced vs. massed distribution instruction for L2 grammar learning. *System*, **42**, 412–428.

Nakata, T. (2015). Effects of expanding and equal spacing on second language vocabulary learning does gradually increasing spacing increase vocabulary learning? *Studies in Second Language Acquisition*, **37**(4), 677–711.

Nakata, T. (2016). Effects of retrieval formats on second language vocabulary learning. *International Review of Applied Linguistics in Language Teaching*, **54**(3), 257–289.

Nakata, T., & Elgort, I. (2021). Effects of spacing on contextual vocabulary learning: Spacing facilitates the acquisition of explicit, but not tacit, vocabulary knowledge. *Second Language Research*, **37**(2), 233–260.

Nakata, T., & Suzuki, Y. (2019). Effects of massing and spacing on the learning of semantically related and unrelated words. *Studies in Second Language Acquisition*, **41**(2), 287–311.

Nakata, T., Tada, S., Mclean, S., & Kim, Y. A. (2021). Effects of distributed retrieval practice over a semester: Cumulative tests as a way to facilitate second language vocabulary learning. *TESOL Quarterly*, **55**(1), 248–270.

Nakata, T., & Webb, S. (2016). Does studying vocabulary in smaller sets increase learning?: The effects of part and whole learning on second language vocabulary acquisition. *Studies in Second Language Acquisition*, **38**(3), 523–552.

Pashler, H., Zarow, G., & Triplett, B. (2003). Is temporal spacing of tests helpful even when it inflates error rates? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **29**(6), 1051–1057.

Pavlik Jr, P. I., & Anderson, J. R. (2005). Practice and forgetting effects on vocabulary memory: An activation-based model of the spacing effect. *Cognitive Science*, **29**(4), 559–586.

Peterson, L. R., Wampler, R., Kirkpatrick, M., & Saltzman, D. (1963). Effect of spacing presentations on retention of a paired associate over short intervals. *Journal of Experimental Psychology*, **66**(2), 206–209.

Plonsky, L., & Derrick, D. J. (2016). A meta-analysis of reliability coefficients in second language research. *The Modern Language Journal*, **100**(2), 538–553.

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

Pyc, M. A., & Rawson, K. A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory & Language*, **60**(4), 437–447.

Raaijmakers, J. G. (2003). Spacing and repetition effects in human memory: Application of the SAM model. *Cognitive Science*, **27**(3), 431–452.

Roediger, H.L., III, & Karpicke, J. D. (2006a). Test enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, **17**, 249–255.

Roediger, H.L., III, & Karpicke, J. D. (2006b). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, **1**(3), 181–210.

Rogers, J. (2015). Learning second language syntax under massed and distributed conditions. *TESOL Quarterly*, **49**, 857–866.

Rogers, J., & Cheung, A. (2020a). Input spacing and the learning of L2 vocabulary in a classroom context. *Language Teaching Research*, **24**(5), 616–641.

Rogers, J., & Cheung, A. (2020b). Does it matter when you review?: Input spacing, ecological validity, and the learning of l2 vocabulary. *Studies in Second Language Acquisition*, 1–19. https://doi.org/10.1017/S0272263120000236

Rohrer, D., & Pashler, H. (2007). Increasing retention without increasing study time. *Current Directions in Psychological Science*, **16**(4), 183–186.

Rundus, D. (1971). Analysis of rehearsal processes in free recall. *Journal of Experimental Psychology*, **89**, 63–77.

Runquist, W. N. (1986). Changes in the rate of forgetting produced by recall tests. *Canadian Journal of Psychology*, **40**, 282–289.

Russo, R., & Mammarella, N. (2002). Spacing effects in recognition memory: When meaning matters. *European Journal of Cognitive Psychology*, **14**(1), 49–59.

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science*, **3**(4), 207–218.

Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language Teaching Research*, **12**(3), 329−363.

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

Serrano, R. (2011). The time factor in EFL classroom practice. *Language Learning*, **61**, 117–145.

Serrano, R., & Huang, H. Y. (2018). Learning vocabulary through assisted repeated reading: How much time should there be between repetitions of the same text? *TESOL Quarterly*, **52**(4), 971–994.

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

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

Shaughnessy, J. J., Zimmerman, J., & Underwood, B. J. (1972). Further evidence on the MP-DP effect in free-recall learning. *Journal of Verbal Learning & Verbal Behavior*, **11**(1), 1–12.

Soderstrom, N. C., Kerr, T. K., & Bjork, R. A. (2016). The critical importance of retrieval—and spacing—for learning. *Psychological Science*, **27**(2), 223–230.

Storm, B. C., Bjork, R. A., & Storm, J. C. (2010). Optimizing retrieval as a learning event: When and why expanding retrieval practice enhances long-term retention. *Memory & Cognition*, **38**(2), 244–253.

Suzuki, Y. (2017). The optimal distribution of practice for the acquisition of L2 morphology: A conceptual replication and extension. *Language Learning*, **67**(3), 512–545.

Suzuki, Y., & DeKeyser, R. (2017). Effects of distributed practice on the proceduralization of morphology. *Language Teaching Research*, **21**(2), 166–188.

Suzuki, Y., Nakata, T., & Dekeyser, R. (2019). The desirable difficulty framework as a theoretical foundation for optimizing and researching second language practice. *The Modern Language Journal*, **103**(3), 713–720.

Suzuki, Y., Nakata, T., & Dekeyser, R. (2020). Empirical feasibility of the desirable difficulty framework: Toward more systematic research on L2 practice for broader pedagogical implications. *The Modern Language Journal*, **104**(1), 313–319.

Suzuki, Y., & Sunada, M. (2020). Dynamic interplay between practice type and practice schedule in a second language: The potential and limits of skill transfer and practice schedule. *Studies in Second Language Acquisition*, **42**(1), 169–197.

Thios, S. J., & D'Agostino, P. R. (1976). Effects of repetition as a function of study-phase retrieval. *Journal of Verbal Learning & Verbal Behavior*, **15**(5), 529–536.

Toppino, T. C., & Gracen, T. F. (1985). The lag effect and differential organization theory: Nine failures to replicate. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **11**(1), 185–191.

Tullis, J. G., Benjamin, A. S., & Ross, B. H. (2014). The reminding effect: Presentation of associates enhances memory for related words in a list. *Journal of Experimental Psychology: General*, **143**(4), 1526–1540.

van den Broek, G. S. E., Takashima, A., Segers, E., & Verhoeven, L. (2018). Contextual richness and word learning: Context enhances comprehension but retrieval enhances retention. *Language Learning*, **68**, 546–585.

Verkoeijen, P., & Bouwmeester, S. (2008). Using latent class modeling to detect bimodality in spacing effect data. *Journal of Memory & Language*, **59**, 545–555.

Verkoeijen, P., Rikers, R., & Schmidt, H. (2004). Detrimental influence of contextual change on spacing effects in free recall. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **30**(4), 796–800.

Verkoeijen, P., Rikers, R., & Schmidt, H. (2005). Limitations to the spacing effect: Demonstration of an inverted u-shaped relationship between inter-repetition spacing and free recall. *Experimental Psychology*, **52**(4), 257–263.

Wegener, S., Wang, H., Beyersmann, E., Nation, K., Colenbrander, D., & Castles, A. (2021, May 19). The effects of spacing and massing on children's orthographic learning. https://doi.org/10.31234/osf.io/d8bmv

Wheeler, M. A., & Roediger, H. L., III. (1992). Disparate effects of repeated testing: Reconciling Ballard's (1913) and Bartlett's (1932) results. *Psychological Science*, **3**, 240–245.

White, J., & Turner, C. (2005). Comparing children's oral ability in two ESL programs. *Canadian Modern Language Review*, **61**(4), 491–517.

Whitten, W. B., & Bjork, R. A. (1977). Learning from tests: Effects of spacing. *Journal of Verbal Learning & Verbal Behavior*, **16**, 465–478.

Young, J. L. (1971). Reinforcement-test intervals in paired-associate learning. *Journal of Mathematical Psychology*, **8**(1), 58–81.

Zechmeister, E. B., & Shaughnessy, J. J. (1980). When you know that you know and when you think that you know but you don't. *Bulletin of the Psychonomic Society*, **15**(1), 41–44.

Zimmerman, J. (1975). Free recall after self-paced study: A test of the attention explanation of the spacing effect. *American Journal of Psychology*, **88**, 277–291.

APPENDIX A

Target Finnish words with their English translations

rakennus	=	building	laukku	=	bag
lehtien	=	leaf	hedelma	=	fruit
sulka	=	feather	perhonen	=	butterfly
sanky	=	bed	ilma	=	air
solmio	=	tie	silta	=	bridge
muna	=	egg	pyrsto	=	tail
pusero	=	shirt	kasine	=	glove
vasara	=	hammer	lapsi	=	child
ruoka	=	food	koira	=	dog
sormi	=	finger	nuoli	=	arrow
lelu	=	toy	piha	=	yard
maaseutu	=	village	hajuvesi	=	perfume
verho	=	curtain	taivas	=	sky
avain	=	key	opettaja	=	teacher
taskuun	=	pocket	kaupunki	=	town
lahja	=	gift	leipa	=	bread
lippu	=	flag	poyta	=	table
orja	=	worker	tarina	=	story
kyna	=	pen	lehma	=	cow
hammas	=	tooth	pilvi	=	cloud
hiekka	=	sand	aurinko	=	sun
keitto	=	soup	jyva	=	grain
ajoneuvo	=	car	veli	=	brother
toimisto	=	office	suihku	=	shower
savuke	=	cigarette	mehu	=	juice
lumi	=	snow	kirjasto	=	library
katu	=	street	kurpitsa	=	pumpkin
kengat	=	shoe	huivi	=	scarf
omena	=	apple	lintu	=	bird
siipi	=	wing	nainen	=	woman
parveke	=	balcony	veitsi	=	knife
kalastaa	=	fish	pelia	=	game
metsa	=	forest	norsu	=	elephant
mekko	=	dress	lusikka	=	spoon
lompakko	=	wallet	lattia	=	floor
tehdas	=	factory	kuva	=	picture

https://doi.org/10.1017/S0142716421000370 Published online by Cambridge University Press

APPENDIX B

Information on the English translations

Frequency and concreteness indices for the English translations for the target words

		English translations								
		LOG10 frequency			CELEX frequency			Concreteness		
List	ISI sublist	M	SD	Mdn	M	SD	Mdn	M	SD	Mdn
A										
	1	1.69	0.50	1.63	84.46	90.65	41.15	4.89	0.11	4.92
	2	1.64	0.42	1.54	72.22	90.96	33.74	4.81	0.14	4.86
	3	1.62	0.46	1.65	67.70	74.53	44.25	4.85	0.11	4.86
	Total for list A	1.65	0.45	1.59	74.79	85.38	39.71	4.85	0.12	4.89
B										
	1	1.56	0.62	1.67	85.14	126.82	46.46	4.81	0.24	4.88
	2	1.85	0.37	1.90	93.59	67.32	78.21	4.60	0.45	4.72
	3	1.57	0.61	1.61	78.49	98.67	39.47	4.81	0.20	4.90
	Total for list B	1.66	0.55	1.75	85.74	97.60	54.71	4.74	0.32	4.84

1 = massed, 2 = short-spaced, 3 = long-spaced

APPENDIX C

Posttests

The form recognition test (Posttest 1)

Please underline the Finnish words that you recognize as ones you have studied during the study phase in this experiment:

varastoi	selkea	kalastaa	tiedon
kyvyn	pysaytti	liittyvan	sotkua
mennyt	nainen	saapui	sotilas
kaytava	piirteensa	ajatteli	koira
kyna	tarina	solmio	keitto
.....	.....	.....	.....
.....	.....	.....	.....

https://doi.org/10.1017/S0142716421000370 Published online by Cambridge University Press

APPENDIX C (*Continued*)
The translation test (Posttest 2)

Please write the English translation next to each Finnish word below:

muna	suihku
sulka	silta
siipi	kirjasto
pilvi	hiekka
leipa	piha
ajoneuvo	omena
.....	.....
.....	.....

The translation matching test (Posttest 3)

Please write the number of each of the English translations below next to its Finnish word on the previous sheet if you were unable to produce its translation from memory:

1. dress	22. balcony
2. fruit	23. brother
3. street	24. food
4. sand	25. table
5. knife	26. village
6. arrow	27. town
7. apple	28. grain
8. game	29. car
.....	.....
.....	.....

Cite this article: Koval, NG. (2022). Testing the reminding account of the lag effect in L2 vocabulary learning. *Applied Psycholinguistics* **43**, 1–40. https://doi.org/10.1017/S0142716421000370