A state-of-the-art review of distribution-of-practice effects on L2 learning

Raquel Serrano
University of Barcelona, Spain
https://orcid.org/0000-0001-9335-4702
raquelserrano@ub.edu

Abstract

The purpose of this state-of-the-art review is to provide a general overview of recent research on time distribution and second language (L2) learning with special implications for classroom settings. Several studies have been performed to examine how to best distribute the hours of L2 practice to maximize learning by comparing conditions that promote intensive exposure versus others in which L2 input or instruction is more widely spaced. Findings from these studies are relevant not only for practical purposes but also for theory development. This review provides a summary of recent studies as well as suggestions for pedagogical practice. Additionally, it identifies areas for future research concerning the effect of time distribution on L2 learning.

Keywords: time distribution; spacing; intensive instruction; distributed practice

1. Introduction

According to DeKeyser (2017), one of the key issues that needs to be addressed in instructed second language acquisition (ISLA) is how to distribute the available instructional time to promote high levels of L2 proficiency. This applies to the school curriculum as well as L2 learning programs for adults in higher education. The topic of time distribution, or input spacing, has attracted the attention

of cognitive psychologists for many years (see Ebbinghaus, 1885/1913) and research on this topic has important pedagogical implications, apart from having an unquestionable theoretical value. In SLA, publications examining the effect of time distribution have drastically increased in the last few years, especially among researchers interested in the role of L2 practice in ISLA (e.g., DeKeyser, 2017; Suzuki et al., 2019).

The *spacing effect*, according to which learning is optimized when repetitions of target material are spaced rather than massed, is one of the most robust findings in cognitive psychology. The evidence shows that including time or other intervening items between repetitions of target items (e.g., *target-distractor-distractor-distractor-target*, etc.) facilitates learning more than subsequent repetitions (e.g., *target-target-target*). The positive effect of spaced as opposed to massed schedules has been found on a variety of tasks and for different population types, even though a lot of research has focused on verbal learning in the case of university students. A related phenomenon is that of the *lag effect*, which suggests that longer intervals between repetitions are more beneficial than shorter ones. The results of previous studies in cognitive psychology indicate that, while the spacing effect is ubiquitous, the lag effect is less consistent (Toppino & Gerbier, 2014).

A recent meta-analysis that focuses on the effects of spacing on L2 learning by Kim and Webb (2022) examines quantitatively the effects of spacing as reported in 37 experimental studies and further confirms the positive role of spacing in general, but points to different effects for different types of L2 areas, learners, or practice activities. Although some of the findings are inconclusive due to the low number of studies and participants in some of the analyses, the meta-analysis confirms the important role of spacing in L2 practice, as well as the need to conduct more research in the area. The present paper complements Kim and Webb's (2022) quantitative meta-analysis in presenting a qualitative narrative review of studies dealing with distribution-of-practice effects and considers not only experimental studies on the learning of a specific target feature, but also studies that have a broader aim and examine the role of spacing at the program level. The current review will provide details on 47 studies on the topic as well as a comprehensive picture of how the distribution of instructional hours has been shown to affect L2 learning.

This review is organized in the following way. Section 2 includes experimental studies and it is further subdivided into four sections: the spacing effect, the lag effect, blocked versus interleaved practice, and individual differences. The next section reviews research at the program level. Section 4 provides a summary of research findings as well as pedagogical implications. The paper concludes with ideas for further research. The Appendix contains details about each of the studies under review (marked with * in the reference list), which will

be useful for the reader, as, due to space limitations, full details cannot be provided for all the studies in the main text.

2. Experimental studies

2.1. Spacing effect

This section reviews the studies that have examined massed versus spaced inter-stimulus spacing when learning occurs in one session, as well as studies that have analyzed learning outcomes of training/teaching in one (massed) versus several sessions (spaced). The first part includes studies on L2 vocabulary, while the second one concerns grammar learning.

2.1.1. Vocabulary

Most research on the spacing effect comes from the cognitive psychology literature, in which, typically, psychology students acquire new vocabulary through paired-associate learning. There are also some studies aiming to contribute to the SLA literature and typically targeting L2 learners, which also use the same methodology. Many of these studies were performed with Japanese English as a foreign language (EFL) university students by Nakata and colleagues and their findings also confirm the spacing effect. Nakata (2015) compared different types of inter-stimulus spacing for the learning of English-Japanese word pairs repeated five times. The results showed that immediate repetitions (massed) promoted less learning than spaced repetitions. Nakata and Suzuki (2019a) provided further support for the spacing effect with English-Japanese translation pairs, which were included in sets of semantically related (e.g., *baboon, badger, otter,* etc.) and semantically unrelated words (e.g., *alcove, pail, pigment*, etc.). The results showed that, although massed repetitions of sets facilitated performance during training, the massed distribution led to significantly fewer vocabulary gains than the spaced one, both on the immediate and delayed posttests.

Further evidence for the spacing effect was provided by Nakata and Elgort (2021) regarding contextual word learning of pseudo-words inserted in English sentences. When the repetition of target items appeared in immediate succession, the participants' performance on the vocabulary tests was worse than when the repetitions were spaced. Interestingly, the authors did not find any differences between conditions in a semantic priming task, supposedly assessing tacit vocabulary knowledge. Koval (2019) used eye-tracking to examine English speakers' processing of Finnish words appearing in English sentences that were repeated consecutively (massed) or with 25 intervening sentences in

between (spaced). The results showed significantly better vocabulary learning results for the spaced condition, for which the decrease in attention as shown by participants' eye movements was not so drastic. In a later study involving Finnish-English paired associates, Koval (2022) found additional evidence for the spacing effect. Her results also showed that massed practice was not significantly different from the no-practice control condition for long-term learning.

Finally, findings from classroom-based studies on L2 English vocabulary learning by L1 Farsi students in primary school suggest that vocabulary practice over two sessions is more conducive to vocabulary learning than one single "massed" session (Lotfolahi & Salehi, 2017).

Apart from the above-mentioned studies focusing on single words, there is research including multi-word units which also provides evidence for the spacing effect. Yamagata et al. (2022) found that spaced repetitions of verb-noun collocations led to more learning not only of the practiced collocations, but also of other collocations with the same target nodes. Similarly, Macis et al. (2021) found a significant advantage for spaced over massed practice for adjective-noun collocations when the training involved deliberate learning. However, the authors found a significant advantage for massed practice in the case of incidental learning.

2.1.2. Grammar

Although there is very little research comparing massed versus spaced grammar instruction, the existing evidence suggests that it is better to use distributed rather than massed practice for long-term learning. Miles (2014) compared massed and widely spaced (average spacing of 2.5 weeks) practice of challenging English grammar structures for Korean university students. The treatment included different classroom activities and the testing consisted of a grammaticality judgment test (GJT) and an L1-L2 translation task. The results showed no differences between conditions on an immediate posttest. On a delayed posttest, however, the spaced group outperformed the massed group on the GJT.

2.2. Lag effects

This section includes the studies that have compared short versus long inter-stimulus spacing in one-session experiments, as well as those analyzing inter-session spacing where learning is distributed over two or more sessions. As Rogers (2017) claims, research on inter-session lags is more relevant for SLA, as learning L2 features typically requires more than one session. The first sub-section focuses on vocabulary, while the next two review the findings concerning grammar learning and speech production.

2.2.1. Vocabulary

While the findings from studies examining the spacing effect in the case of vocabulary learning are quite consistent, the evidence regarding the lag effect is not so uniform. Nakata (2015) found no differences between short (5 items), medium (10 items) and long (30 items) inter-stimulus spacing, while Nakata and Webb (2016) reported that longer inter-repetition lags of 19 items were more beneficial for long-term learning of vocabulary than shorter lags of 3 items. Koval (2022) found that long spacing (71-119 trials within a block) was more beneficial than short spacing (17-38) for long-term learning of Finnish-English paired associates. As can be observed, "short" and "long" spacing were differently operationalized, which might partly explain the inconsistent results.

On the other hand, research on inter-session spacing in classroom settings has generally not provided support for the lag effect. Küpper-Tetzel et al. (2014a) compared the learning of German-English word pairs by German grade 6 learners under a massed schedule and two spaced schedules (1-day and 10-day lags). The results of a 7-day delayed test showed that the 1-day lag was more beneficial than the other two. Five weeks later, the two spaced conditions proved more advantageous than the massed condition, with no significant differences between the short and long lags. These results support previous claims that the optimal inter-session interval depends on the retention interval (Cepeda et al., 2006).

Rogers and Cheung (2020, 2021) examined lag effects for vocabulary learning in a primary school in Hong Kong. In the first study, the target words that were learned over a short 1-day lag were better remembered 28 days later than those learned over a longer 8-day lag. However, the second study, which was a replication of the first, found no differences between lags.

In contrast to the previous studies examining the learning of L2 words in isolation, the studies by Serrano and Huang (2018, 2021) focused on contextual word learning through repeated reading in the case of secondary-school students in Taiwan. The results similarly failed to provide support for the lag effect. In both studies, the intensive condition (1-day inter-session interval) led to higher vocabulary gains on the immediate posttest than the long-spaced condition (7 days). Performance on the delayed posttest differed when learning was incidental, with no differences between conditions (Serrano & Huang, 2018), or intentional, in which case higher gains were reported for the intensive condition (Serrano & Huang, 2021).

Finally, there are some studies that have analyzed whether changing the intervals between lags during the treatment is more or less beneficial than equal or uniform spacing. Nakata (2015) found an advantage of expanding (gradually increasing inter-repetition intervals) over equal spacing in learning performed in one session. Studies examining learning over multiple sessions have reported conflicting results. For

example, Küpper-Tetzel et al. (2014b) found no significant differences between contracting (from 5-day to 1-day lags), equal (3-day lag) and expanding (from 1- to 5-day lags) on an immediate posttest. On a test performed 7 days later, the contracting schedule was better than the equal and the expanding ones, while the opposite was found 35 days after training, with the equal and expanding schedules outperforming the contracting schedule. These results contrast with the findings from Schuetze and Weimer-Stuckmann (2011), which showed no differences between equal and expanding schedules for short-term learning but better retention in the uniform condition. In another study comparing uniform and expanding schedules for the learning of English-German word pairs, Schuetze (2015) did not find any significant differences between the two. Similarly, Snoder (2017) did not find any significant differences in the learning of verb-noun collocations between an expanding schedule (day 1, 7, and 16) and an intensive schedule (day 1, 2, 4).

2.2.2. Grammar

The first studies examining the effect of inter-session spacing in SLA concerned grammar learning in classroom settings and provided support for the lag effect. Bird (2010) focused on the acquisition of the simple past, present perfect and past perfect by adult EFL learners in Malaysia, over five different class sessions, spaced either over a 3-day or a 14-day interval. The results of a 7-day delayed GJT showed no differences between groups. However, the longer lag proved more helpful for long-term retention after 60 days. Rogers (2015) provided further evidence for the benefit of spacing grammar instruction over longer lags (2.25 vs. 7 days) for the incidental acquisition of challenging English grammar structures by a group of university students in the Middle East.

In contrast, the study by Kasprowicz et al. (2019), which examined the acquisition of French morphology by L1 English learners of French in grades 4-6, did not find any differences between short (3.5 days) and long (7 days) lags either on an immediate or delayed posttest.

Research by Suzuki and colleagues on productive grammar skills also failed to support the lag effect. What is more, their findings suggested that short lags might be more beneficial than longer lags. Suzuki and DeKeyser (2017a) compared 1- versus 7-day inter-session intervals for short- and long-term learning of Japanese morphology by adult English speakers. Learning was assessed through accuracy and speed of performance in a rule application and a sentence completion test. The results showed no differences between lags for accuracy; however, the short-lag condition led to significantly faster performance 28 days after the instruction. In a conceptual replication and extension of that study, Suzuki (2017) provided more evidence in favor of short lags (3.3 vs. 7 days), but concerning accuracy and not speed in the production

of morphology in a novel miniature language (Supurango) by L1 Japanese university students. In a follow-up study (Suzuki, 2018), it was found that the short-lag condition was also more conducive to automatization, as evidenced by participants' scores in the CV (coefficient of variation) (Segalowitz & Segalowitz, 1993).

2.2.3. L2 speech production

Two different aspects of students' L2 speech production have been examined regarding lag effects, one being pronunciation and the other oral fluency. Li and DeKeyser (2019) examined the acquisition of tonal word production in Chinese. The training involved the presentation of target words as well as practice that was meant to promote different types of knowledge: declarative (knowing "what," such as knowing about different tones in Mandarin) and procedural (knowing "how;" e.g., how to use the right tone in oral speech production). The authors reported that declarative knowledge decreased significantly when tested 28 days later. In addition, when the lags between training sessions were longer (7 days), this declarative knowledge was better retained than when there was only a 1-day lag. However, it was observed that for the production of new words, involving procedural knowledge, short spacing was more beneficial.

In the case of oral fluency, Bui et al. (2019) examined the effect of task repetition under different schedules for the development of L2 oral complexity, accuracy, and fluency. The same task was repeated twice either immediately (massed) or 1, 3, 7, or 15 days after the first performance. Whereas no differences were found in terms of complexity and accuracy, immediate task repetition led to significantly higher fluency than its spaced counterpart, while no other differences were found between other lags.

In a more thorough investigation of oral fluency, Suzuki and Hanzawa (2022) examined the effect of spacing six repetitions of the same task, and compared massed (immediate) short (45 minutes) and long (7 days) spacing. The authors found massed repetitions to be a "double-edged sword," because they were helpful in significantly reducing students' pauses but also led to slower articulation rate and more verbatim repetitions.

2.3. Blocking versus interleaving

Also related to time distribution, some other studies in the SLA literature have focused on whether it is more effective to learn similar forms in blocks, in which repetitions of target items or examples of target rules appear subsequently (i.e., massed), or whether interleaved practice (alternating between repetition types, i.e., spaced) is more beneficial for L2 learning.

Nakata and Suzuki (2019b) examined the learning of three categories: English simple past, present perfect, and the conditionals by Japanese university students. Under the blocked condition, the activity included structures from each category consecutively. The interleaved condition alternated sentences from different categories, while in the increasing condition five sentences from each category were practiced first in blocks, while the other five were interleaved. The results of an immediate GJT showed no differences between conditions; however, the results of a delayed posttest 7 days later were significantly higher under interleaved than under blocked practice.

Suzuki and Sunada (2020) also compared these three types of schedules but, in contrast to the previous study, found the hybrid schedule (first blocked and then interleaved) to be more beneficial for the acquisition of relative pronouns. In another study also examining the learning of English relative clauses by Japanese learners but only under two schedules (blocked vs. interleaved), Suzuki et al. (2022b) showed that interleaving was more helpful for fast and accurate oral production of relative clauses on an immediate posttest, while no differences between conditions were found on a 7-day delayed posttest.

In the case of oral fluency, Suzuki (2021) found that repeating the same task three times in blocks (AAA BBB CCC) led to more fluent speech than interleaving different tasks (ABC ABC ABC). Additionally, the learners doing blocked practice were more likely to reuse the same constructions (Suzuki et al., 2022a).

Carpenter and Mueller (2013) also compared blocked and interleaved practice for the learning of eight French-pronunciation rules by L1 English speakers. The authors found that blocking (presenting example words for each rule subsequently – *bateau*, *carreau*, *fardeau*, etc.) was more helpful for learning pronunciation than interleaved practice, in which the presentation sequence alternated words following different rules (*bateau, genou, tandis*, etc.).

2.4. Lag effects and individual differences

Several studies have investigated whether certain cognitive capacities differentially affect learning under more or less concentrated schedules. Most of this research has been done by Suzuki and colleagues within the aptitude-treatment interaction framework (Robinson, 2002).

Suzuki (2018) examined the role of procedural learning ability, related to the acquisition of fast and automatized knowledge, and found that it plays a clearer role when learning L2 grammar under short (3.3 days) rather than long (7 days) inter-session lags. Several studies have focused on the role of working memory (WM), which refers to a limited-capacity complex cognitive system that allows for the storage and processing of information while performing cognitive tasks (Baddeley, 2003). Different instruments have been used to measure WM; for instance, Suzuki and DeKeyser

(2017b), and Suzuki (2019) used an operation span task; Suzuki (2021a) a trail-making task, while Suzuki et al. (2022a) measured WM through a listening span task. The role of WM in learning under different schedules is still unclear, although most of the evidence suggests that WM plays a more notable role in learning under concentrated schedules. Suzuki and DeKeyser (2017b) found that WM predicted learning of Japanese morphosyntax when inter-session spacing included short (1 day) but not long (7 days) lags. In a similar vein, learners' WM has been shown to affect their oral fluency development (Suzuki, 2021a), as well as their learning of relative clauses (Suzuki et al., 2022b) under blocked but not under interleaved practice schedules. In contrast, Suzuki (2019) found no effect of WM for the learning of Supurango under short (3.3 days) versus long (7 days) lags, even though the study also included an operation span task as in Suzuki and DeKeyser (2017b).

As for language-analytic or grammar-inferencing abilities, Suzuki and DeKeyser (2017b) and Suzuki (2019) reported that these skills had a clearer role when participants were learning L2 grammar in long-spaced sessions (7 days). Under this type of spaced schedule, the participants who were better able to infer grammar rules in an unknown language or memorize new form-meaning mappings (as measured by LLAMA-F and LLAMA-B, Meara, 2005) were more successful in learning the target L2 grammar.

Using the desirable difficulties framework (Bjork, 1994; Suzuki et al., 2019), Serfaty and Serrano (2022) examined how learners' individual characteristics regarding language proficiency, age, and time on task during training predicted grammar learning through digital flashcards. The authors found no overall lag effects (1-day vs. 7-day lags) when the data from all the students were analyzed together, but crucially, their analyses showed that the longer lag was more beneficial for learners of higher proficiency and shorter times on task during the learning phase, while the opposite was true for learners experiencing more difficulty during training. In other words, the longer lag was a desirable difficulty only when no additional difficulties existed on the part of the learner.

3. Spacing and program evaluation

The final set of studies in this review includes those focusing on the effect of time distribution at the program level, comparing programs in which the hours of instruction were differently distributed. Research in this area is scarce, with many of these studies being performed in primary schools in Canada, where a change was implemented in the 1980s to promote intensive English instruction in Quebec.

In order to extend the findings from an earlier large-scale study involving thousands of students in Quebec by Spada and Lightbown (1989) showing significant advantages in favor of learners receiving intensive English instruction, White

and Turner (2005) performed an exhaustive analysis of students' oral production. This study compared the oral performance of learners receiving intensive (400 hrs in one year) versus regular (±60 h) instruction on a variety of oral tasks. Their results showed that the oral communicative abilities of the learners in the intensive program were significantly more advanced than their peers' receiving regular instruction. More recently, French et al. (2020) examined the long-term effects of intensive instruction on speech production in terms of perceived fluency, comprehensibility and accentedness. The authors found that four years after the end of their respective programs the students that had been enrolled in intensive English were perceived to be more fluent and comprehensible in this language than those who had only received regular instruction. No differences were found in accentedness, according to the raters' perceptions. The authors controlled for students' academic and language skills and, although there might be other intervening variables that were not controlled for, the results of this study provide evidence for the positive effect of intensive instruction.

Other studies were also performed in Canada in which the amount of exposure was held constant, focusing on different implementations of intensive English, referred to as massed (300-400 hrs over five months) and distributed (same hours over ten months). Collins et al. (1999) compared the learning outcomes of a group of students (N = 700) enrolled in these two programs as well as in a massed plus program, which promoted out-of-class L2 use in the school. The students performed different tests that tapped different L2 skills at the end of their respective program, which showed that the learners in the massed programs significantly outperformed those in the distributed program in most measures. However, the authors caution about attributing the difference exclusively to the distribution of instructional hours, as the students in the massed programs also ended up receiving a few more hours of instruction. Collins and White (2011) replicated these results. The authors performed a longitudinal study and assessed learners' L2 skills at four different 100-hour intervals. Although the authors suggest that the differences were not large and some of them might be due to instructional practice, several statistically significant differences were found, especially at time 3 and 4, in favor of the concentrated program.

In the Spanish context, studies by Serrano and colleagues (Serrano, 2011; Serrano & Muñoz, 2007; Serrano et al., 2015) analyzed L2 development in English courses that offered the same number of hours of instruction but distributed differently (110 hrs in 1 month vs. 3-4 vs. 7 months) in the case of adult EFL learners in a university setting. Apart from performing a general proficiency test, the participants did an oral narrative and a written essay before and after their respective course. The results showed some advantages to the more intensive program, but only at

the beginner or intermediate level and for a few measures, mostly related to grammar and lexical richness and use of formulaic language in oral production.

Alcaraz-Mármol (2015) examined vocabulary learning after a 2-month intensive (6 hrs/week) and 6-month extensive (2 hrs/week) course, also in the case of adult Spanish EFL learners. The intensive program promoted more significant vocabulary gains and, although the learners also experienced more losses on a delayed posttest 10 weeks later, their performance was still significantly superior to their peers' in the extensive program.

These results contrast with the findings from Xu et al. (2012) for a group of high school learners of Mandarin in the US. In this study, although most comparisons showed no difference between a summer intensive program and a semester-long program offering the same number of hours of instruction, the learners in the latter program became more fluent.

4. Summary of findings and implications for L2 teaching

As can be seen from the overview presented above, the results of the studies conducted so far present some conflicting evidence for the role of spacing in different areas of L2 learning in experimental studies. The findings from these studies also contrast with those analyzing the role of intensity at the program level.

Concerning experimental studies, there is one robust finding: when learning vocabulary items from lists in one session (either including L2-L1 pairs or in sentences), it is better to space repetitions than studying them in massed sequences (e.g., Koval, 2019, 2022; Nakata, 2015; Nakata & Suzuki, 2019a). Considering this finding, L2 learners should not engage in repetitive blocked/massed practice of each individual item when they are revising/learning new vocabulary from lists, but instead go through the whole list before doing repeated practice of individual words.

The results comparing vocabulary learning in one session versus several sessions show better learning outcomes under the latter schedule. This evidence suggests that teachers should encourage their learners to revise their vocabulary periodically on different days and not just one day before a test. As Nakata et al. (2021) suggest, cumulative testing might be a good way to promote vocabulary learning over different sessions, at the same time as it increases the amount of learning opportunities. However, it is not clear yet how long inter-session lags should be in spaced vocabulary practice, as some studies have found an advantage to shorter lags (Rogers & Cheung, 2020; Serrano & Huang, 2021) and others have found little difference (Rogers & Cheung, 2021). There is some indication, however, that longer lags might be more favorable when knowledge is assessed after a long period, suggesting, again, that spacing repeated exposures to novel words in the classroom is positive if long-term knowledge is the goal (Küpper-Tetzel et al., 2022a).

If we now turn to grammar learning, the results comparing massed and spaced schedules go in the same direction as for vocabulary. Interleaved or hybrid grammar practice, in which exemplars of target rules do not appear subsequently but are interspersed, promote better long-term results than blocked practice, with learning taking place in one session (Nakata & Suzuki, 2019b; Suzuki & Sunada, 2020; Suzuki et al., 2022b). One pedagogical recommendation following these findings would be for teachers to focus on contrasting different structures in one session (for instance, simple past and present perfect), rather than devoting the whole session to one single structure. Similarly, following Miles (2014), it is advisable to devote more than one session to the teaching of L2 grammar forms, which probably represents typical classroom practice in most contexts.

Concerning the lag effect in grammar learning over multiple sessions, there is conflicting evidence. On the one hand, some classroom-based studies support the lag effect for long-term learning, mostly for receptive grammar knowledge assessed through GJTs (Bird, 2010; Rogers, 2015), while, on the other hand, experimental studies examining productive skills either report no differences between lags (Serfaty & Serrano, 2022) or an advantage to shorter lags (Suzuki, 2017; Suzuki & DeKeyser, 2017a). One teaching implication would be that, if receptive declarative knowledge is the goal, it might be better to include longer lags between practice sessions, while for the proceduralization of grammar rules, shorter lags might be more beneficial.

The conflicting results obtained for lag effects for grammar might be due to the type of training and testing (receptive vs. productive skills) used in the different studies, or to learners' individual differences that were not controlled for. As some studies have shown (see section 2.4), certain types of learner profiles might benefit more from shorter or longer lags. According to Suzuki et al. (2019) and as shown in research by Serfaty and Serrano (2022), longer lags are a source of difficulty that might not be desirable when there are additional sources of difficulty on the part of the learner (e.g., low proficiency or challenges during the learning phase). It might be advisable for teachers to consider the characteristics of their learners when deciding how to space grammar practice and include longer lags in advanced groups and shorter when the group's proficiency is low. However, adapting to individual learners within a group might be challenging in classes where learners' characteristics are very diverse.

Research on L2 speech fluency suggests that massed or blocked practice could be more beneficial for the proceduralization of oral production skills (Bui et al., 2019; Suzuki, 2021b). However, when there are too many repetitions, massed practice might no longer be optimal (Suzuki & Hanzawa, 2022). Regarding pronunciation rules, the evidence suggests that more concentrated practice (blocked, if done in one session or under short lags if done over several sessions)

might be more helpful for learning L2 pronunciation (Carpenter & Mueller, 2013; Li & DeKeyser, 2019). According to these findings, L2 classes should offer students the possibility of repeating oral fluency tasks or doing repeated productive or receptive practice of pronunciation rules under short-spaced schedules.

At the program level, the findings from the Canadian studies on intensive English in primary education provide clear support for intensive instruction, especially when it involves more contact hours (e.g., White & Turner, 2005) or when it is concentrated under shorter time periods. Although the differences between concentrated versus distributed intensive programs are not large, when they exist, they are in favor of the more concentrated schedule (Collins et al., 1999; Collins & White, 2011). The results of the comparison between intensive versus regular programs for adult learners are not conclusive, but, along the same lines, there appear to be more advantages for intensive L2 learning (e.g., Serrano, 2011). As Lightbown (2014) claims, drip-feed L2 instruction including very few hours per week (often 1-2), which is the typical schedule in most educational contexts, does not lead to advanced L2 skills or high communicative competence (Stern, 1985). Instead, L2 programs should offer full-flow (or intensive) exposure to the target language (see also Muñoz, 2012). Lightbown and Spada (2020) claim that it is beneficial to concentrate L2 instructional hours at the curricular level, even when there is no time increase. The authors suggest that increasing and concentrating the amount of L2 instruction when the students are more cognitively mature results in better L2 learning outcomes than an earlier start. The provision of intensive English in schools in Quebec required some restructuring of schedules for other subjects in the school curriculum, which might be challenging in many contexts. However, the promising results obtained in Canada could encourage the implementation of equivalent programs in other contexts.

It must be emphasized that the way "intensity" is conceptualized at the program level is different from the experimental studies, as it refers to intensity of total time devoted to L2 learning and not the (repetitive) practice of a specific target form (see Serrano, 2012).

5. Conclusion and further research

The results of the studies included in this review show that the findings on the spacing effect from cognitive psychology apply to the SLA literature for learning that takes place under similar conditions, typically rote learning of L2 vocabulary from lists. For the development of declarative knowledge (e.g., knowledge about rules), it is more beneficial to learn/practice in more than one session. It is not clear, however, whether adding more space between learning sessions is always more beneficial for L2 learning, and, in some cases, there is evidence to the

contrary, as in the case of fluent (or proceduralized) L2 production, or when learning difficulties exist on the part of the learner. Further research is needed in order to know more about what L2 areas might benefit from longer spacing and for what type of learners. Recent studies on individual differences are throwing some new light on the spacing literature; however, this research is still scarce. Considering the program-evaluation literature, long spacing of small "L2 doses" is probably not recommended. It seems reasonable to assume that L2 learning, as L1 learning, should also require high doses of the target language or a full-flow approach (Lightbown, 2014; Stern, 1985). However, we need more studies that investigate how time distribution affects learning at the program level, or the development of general L2 skills both in the case of children and adults, especially considering long-term retention, which has been under-analyzed in previous research. Moreover, future research at the program level should control more the actual teaching practice, although this might be challenging considering the amount of hours of instruction that are usually involved in this type of research.

One point that needs to be mentioned is that, with the exception of the Canadian programs, most studies in this review have analyzed data from small samples and, in some cases, only around 15 learners in some conditions. These small sample sizes might be responsible for the conflicting results that are sometimes reported. Although gathering data from large samples is always a challenge in SLA research, future studies should try to obtain data from larger groups.

There are currently some replication studies (e.g., Rogers & Cheung, 2021; Serrano & Huang, 2021; Suzuki, 2017); however, more replication or close replication studies would be desirable to check whether previous findings are generalizable to other participants under equivalent methodological conditions or to different age groups. While there are some experimental studies with primary or secondary school students, most studies target adults. Additionally, areas other than vocabulary and grammar should be given priority in future studies, as most of the evidence we now have comes from research examining these two areas. More research is also necessary investigating different types of knowledge in the same study to confirm previous claims that they might be differentially affected by spacing (e.g., declarative vs. procedural; intentional vs. incidental, etc.) Finally, although it is important to have information about learning outcomes under different practice schedules, more research should be performed also examining learning processes, for instance by using eye-tracking (as in Koval, 2019) or analyzing learners' performance during the learning phase (e.g., Nakata & Suzuki, 2019a), as this research throws more light on how spacing affects L2 learning and also contributes to theoretical explanations of the spacing/lag effects.

In summary, the findings reported in this state-of-the-art review point to the need for more research on the effect of time distribution. The field needs

more conclusive evidence in order to offer both practitioners and policy makers concrete and scientifically supported advice about how to organize the often limited available time for L2 learning.

Acknowledgements

References[1]

*Alcaraz-Mármol, G. (2015). Second language vocabulary acquisition under two different types of instruction: The effects of concentrated and distributed introduction in immediate and delayed retention. *ES Revista De Filologia Inglesa*, *36*, 7-25. https://core.ac.uk/reader/211105905

Baddeley, A. (2003). Working memory and language: An overview. *Journal of Communication Disorders, 36*, 189-208. https://doi.org/10.1016/S0021-9 924(03)00019-4

*Bird, S. (2010). Effects of distributed practice on the acquisition of second language English syntax. *Applied Psycholinguistics*, *31*(4), 635-650. https://doi.org/10. 1017/s0142716410000172

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.

*Bui, G., Ahmadian, M. J., & Hunter, A-M. (2019). Spacing effects on repeated L2 task performance. *System*, *81*, 1-13. https://doi.org/10.1016/j.system.2018.12.006

*Carpenter, S. K., & Mueller, F. E. (2013). The effects of interleaving versus blocking on foreign language pronunciation learning. *Memory & Cognition*, *41*, 671-682. https://doi.org/10.3758/s13421-012-0291-4

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

*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. https:// doi.org/10.2307/3587881

*Collins, L., & White, J. (2011). An intensive look at intensity and language learning. *TESOL Quarterly*, *45*(1), 106-133. https://doi.org/10.5054/tq.2011.240858

DeKeyser, R. M. (2017). Knowledge and skill in ISLA. In S. Loewen & M. Sato (Eds.). *Routledge handbook of instructed second language acquisition* (pp. 15-32). Routledge.

Ebbinghaus, H. (1913). *Memory: A contribution to experimental psychology*. Teachers College, Columbia University. (Original work published 1885)

*French, L. M., Gagné, N., & Collins, L. (2020). Long-term effects of intensive instruction on fluency, comprehensibility and accentedness. *Journal of Second Language Pronunciation*, *6*(3), 380-401. https://doi.org/10.1075/jslp.20026.fre

*Kasprowicz, R. E., Marsden, E., & Sephton, N. (2019). Investigating distribution of practice effects for the learning of foreign language verb morphology

---

[1] The references marked with * are included in the review.

in the young learner classroom. *Modern Language Journal*, *103*(3), 580-606. https://doi.org/10.1111/modl.12586

Kim, S. K., & Webb, S. (2022). The effects of spaced practice on second language learning: A meta-analysis. *Language Learning.* Advance online publication https://doi.org/10.1111/lang.12479

*Koval, N. G. (2019). Testing the deficient processing account of the spacing effect in second language vocabulary learning: Evidence from eye tracking. *Applied Psycholinguistics*, *40*(05), 1103-1139. https://doi.org/10.1017/s0142716419000158

*Koval, N. G. (2022). Testing the reminding account of the lag effect in L2 vocabulary learning. *Applied Psycholinguistics, 43*(1), 1-40. https://doi.org/10.1017/S0142716421000370

*Küpper-Tetzel, C. E., Erdfelder, E., & Dickhäuser, O. (2014a). The lag effect in secondary school classrooms: Enhancing students' memory for vocabulary. *Instructional Science*, *42*, 373-388. https://doi.org/10.1007/s11251-013-9285-2

*Küpper-Tetzel, C. E., Kapler, I. V., & Wiseheart, M. (2014b). Contracting, equal, and expanding learning schedules: The optimal distribution of learning sessions depends on retention interval. *Memory & Cognition*, *42*(5), 729-741. https://doi.org/10.3758/s13421-014-0394-1

*Li, M., & DeKeyser, R. M. (2019). Distribution of practice effects in the acquisition and retention of L2 Mandarin tonal word production. *Modern Language Journal*, *103*(3), 607-628. https://doi.org/10.1111/modl.12580

Lightbown, P. M. (2014). Making the minutes count in L2 teaching. *Language Awareness, 23*(1-2), 3-23. https://doi.org/10.1080/09658416.2013.863903

Lightbown, P. M., & Spada, N. (2020). Teaching and learning L2 in the classroom: It's about time. *Language Teaching*, *53*(4), 422-432. https://doi.org/10.1017/S0261444819000454

*Lotfolahi, A. R., & Salehi, H. (2017). Spacing effects in vocabulary learning: Young EFL learners in focus. *Cogent Education*, *4*(1), 1287391. https://doi.org/10.1080/2331186x.2017.1287391

*Macis, M., Sonbul, S., & Alharbi, R. (2021). The effect of spacing on incidental and deliberate learning of L2 collocations. *System, 103*, 9102649. https://doi.org/10.1016/j.system.2021.102649

Meara, P. M. (2005). *LLAMA language aptitude tests.* Lognostics.

*Miles, S. W. (2014). Spaced vs. massed distribution instruction for L2 grammar learning. *System*, *42*, 412-428. https://doi.org/10.1016/j.system.2014.01.014

Muñoz, C. (2012). *Intensive exposure experiences in second language learning.* Multilingual Matters.

*Nakata, T. (2015). Effects of expanding and equal spacing on second language vocabulary learning: Does gradually increasing spacing increase vocabulary

learning? *Studies in Second Language Acquisition*, *37*(4), 677-711. https://doi.org/10.1017/s0272263114000825

*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. https://doi.org/10.1177/0267658320927764

*Nakata, T., & Suzuki, Y. (2019a). Effects of massing and spacing on the learning of semantically related and unrelated words. *Studies in Second Language Acquisition*, *41*(2), 287-311. https://doi.org/10.1017/s0272263118000219

*Nakata, T., & Suzuki, Y. (2019b). Mixing grammar exercises facilitates long-term retention: Effects of blocking, interleaving, and increasing practice. *Modern Language Journal*, *103*(3), 629-647. https://doi.org/10.1111/modl.12581

Nakata, T., Tada, S., McLean, S., & Kim, Y. E. (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. https://0.1002/tesq.596

*Nakata, T., & Webb, S. (2016). Does studying vocabulary in smaller sets increase learning? The effects of part and whole learning on second language vocabulary acquisition. *Studies in Second Language Acquisition*, *38*(3), 523-552. https://doi.org/10.1017/s0272263115000236

Robinson, P. (2002). Effects of individual differences in intelligence, aptitude and working memory on adult incidental SLA. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 211-266). John Benjamins.

*Rogers, J. (2015). Learning second language syntax under massed and distributed conditions. *TESOL Quarterly*, *49*(4), 857-866. https://doi.org/10.1002/tesq.252

Rogers, J. (2017). The spacing effect and its relevance to second language acquisition. *Applied Linguistics, 38*(6), 906-911. https://doi.org/10.1093/applin/amw052

*Rogers, J., & Cheung, A. (2020). Input spacing and the learning of L2 vocabulary in a classroom context. *Language Teaching Research*, *24*(5), 616-641. https://doi.org/10.1177/1362168818805251

*Rogers, J., & Cheung, A. (2021). Does it matter when you review? Input spacing, ecological validity, and the learning of L2 vocabulary. *Studies in Second Language Acquisition*, *43*(5), 1138-1156. https://doi.org/10.1017/s0272263120000236

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

*Schuetze, U., & Weimer-Stuckmann, G. (2011). Retention in SLA lexical processing. *CALICO Journal*, *28*(2), 460-472. https://doi.org/10.11139/cj.28.2.460-472

Segalowitz, N. S., & Segalowitz, S. J. (1993). Skilled performance, practice, and the differentiation of speed-up from automatization effects: Evidence

from second language word recognition. *Applied Psycholinguistics, 14*(3), 369-385. https://doi.org/10.1017/S0142716400010845

*Serfaty, J., & Serrano, R. (2022). Lag effects in grammar learning: A desirable difficulties perspective. *Applied Psycholinguistics, 43*(3), 513-550. https://doi.org/10.1017/S0142716421000631

*Serrano, R. (2011). The time factor in EFL classroom practice. *Language Learning, 61*(1), 117-145. https://doi.org/10.1111/j.1467-9922.2010.00591.x

Serrano, R. (2012). Is intensive learning effective? In C. Muñoz (Ed.), *Intensive exposure experiences in second language learning* (pp. 3-22). Multilingual Matters.

*Serrano, R., & Huang, H. Y. (2018). Learning vocabulary through assisted repeated reading: How much time should there be between repetitions of the same text? *TESOL Quarterly, 52*(4), 971-994. https://doi.org/10.1002/TESQ.445

*Serrano, R., & Huang, H. Y. (2021). Time distribution and intentional vocabulary learning through repeated reading: A partial replication and extension. *Language Awareness*. Advance online publication. https://doi.org/10.1080/09658416.2021.1894162

*Serrano, R., & Muñoz, C. (2007). Same hours, different time distribution: Any difference in EFL? *System, 35*(3), 305-321. https://doi.org/10.1016/j.system.2007.02.001

*Serrano, R., Stengers, H., & Housen, A. (2015). Acquisition of formulaic sequences in intensive and regular EFL programs. *Language Teaching Research, 19*(1), 89-106. https://doi.org/10.1177/1362168814541748

*Snoder, P. (2017). Improving English learners' productive collocation knowledge: The effects of involvement load, spacing, and intentionality. *TESL Canada Journal, 34*(3), 140-164. https://doi.org/10.18806/tesl.v34i3.1277

Spada, N., & Lightbown, P. M. (1989). Intensive ESL programs in Quebec primary schools. *TESL Canada Journal, 7*(1), 11-32.

Stern, H. H. (1985). The time factor and compact course development. *TESL Canada Journal, 3*(1), 13-29. https://doi.org/10.18806/tesl.v3i1.471

*Suzuki, Y. (2017). The optimal distribution of practice for the acquisition of L2 morphology: A conceptual replication and extension. *Language Learning, 67*(3), 512-545. https://doi.org/10.1111/lang.12236

*Suzuki, Y. (2018). The role of procedural learning ability in automatization of L2 morphology under different learning schedules. *Studies in Second Language Acquisition, 40*(4), 923-937. https://doi.org/10.1017/s0272263117000249

*Suzuki, Y. (2019). Individualization of practice distribution in second language grammar learning. *Journal of Second Language Studies, 2*(2), 169-196. https://doi.org/10.1075/jsls.18023.suz

*Suzuki, Y. (2021a). Individual differences in memory predict changes in breakdown and repair fluency but not speed fluency: A short-term fluency training

intervention study. *Applied Psycholinguistics, 42*(4), 969-995. https://doi.org/10.1017/S0142716421000187

*Suzuki, Y. (2021b). Optimizing fluency training for speaking skills transfer: Comparing the effects of blocked and interleaved task repetition. *Language Learning, 71*(2), 285-325. https://doi.org/10.1111/LANG.12433

*Suzuki, Y., & DeKeyser, R. M. (2017a). Effects of distributed practice on the proceduralization of morphology. *Language Teaching Research, 21*(2), 166-188. https://doi.org/10.1177/1362168815617334

*Suzuki, Y., & DeKeyser, R. M. (2017b). Exploratory research on second language practice distribution: An Aptitude × Treatment interaction. *Applied Psycholinguistics, 38*(1), 27-56. https://doi.org/10.1017/s0142716416000084

*Suzuki, Y., & Hanzawa, K. (2022). Massed task repetition is a double-edged sword for fluency development: An EFL classroom study. *Studies in Second Language Acquisition, 44*(2), 536-561. https://doi.org/10.1017/S0272263121000358

*Suzuki, Y., & Sunada, M. (2020). Dynamic interplay between practice type and practice schedule in a second language. *Studies in Second Language Acquisition, 42*(1), 169-197. https://doi.org/10.1017/s0272263119000470

*Suzuki, Y., Eguchi, M., & de Jong, N. (2022a). Does the reuse of constructions promote fluency development in task repetition? A usage-based perspective. *TESOL Quarterly.* Advance online publication. https://doi.org/10.1002/tesq.3103

Suzuki, Y., Nakata, T., & DeKeyser, R. M. (2019). The desirable difficulty framework as a theoretical foundation for optimizing and researching second language practice. *Modern Language Journal, 103*(3), 713-720. https://doi.org/10.1111/modl.12585

*Suzuki, Y., Yokosawa, S., Aline, D. (2022b). The role of working memory in blocked and interleaved grammar practice: Proceduralization of L2 syntax. *Language Teaching Research, 26*(4), 671-695. https://doi.org/10.1177/1362168820913985

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

*White, J., & Turner, C. (2005). Comparing children's oral ability in two ESL programs. *Canadian Modern Language Review, 61*(4), 491-517. https://doi.org/10.3138/cmlr.61.4.491

*Xu, X., Padilla, A. M., & Silva, D. (2012). The time factor in Mandarin language learning: The four-week intensive versus the regular high school semester. *The Language Learning Journal, 42*(1), 55-66. https://doi.org/10.1080/09571736.2012.677054

*Yamagata, S., Nakata, T., & Rogers, J. (2022). Effects of distributed practice on the acquisition of verb-noun collocations. *Studies in Second Language Acquisition.* Advance online publication. https://doi.org/10.1017/S0272263122000225

APPENDIX

List of studies included in the review (see list of abbreviations in alphabetical order at the end)

Study	SP LAG B/INT PROG	L2 area	Participants Age, context	Training	Tests/ instruments	N. of sessions/ hours	Spacing	Retention in days (unless specified)	Results
Alcaraz-Mármol (2015)	PROG	Vocabulary (96 words)	60 L1 Spa, L2 English University	Classroom instruction	L1-L2 translation, L2-L1 translation	48 hrs	Int.: 6h/week over 2 months Extensive: 2h/week over 6 months	Immediate RI-10 weeks	Intensive > extensive both immediate and delayed
Bird (2010)	LAG	Grammar (simple past, present perf., past perf.)	38 L1 Malay, L2 English University	Sentence correction exercises	GJT 160 SP-PP 160 PP-PP Test 2 sets of 20	28 hrs 10 sessions	ISI-3 ISI-14	RI-7 RI-60	No sig. diff. at 7-day RI; Distributed > concentrated at 60-day RI
Bui et al. (2019)	SP LAG	Oral complexity, accuracy, fluency (CAF)	71 L1 Cantonese, L2 English University	Oral picture-based task	Oral picture-based task	2 repetitions	Massed ISI-1 ISI-3 ISI-7 ISI-14 days	Immediate	Only diff.: Speech rate: immediate > ISI-1, ISI-7, ISI-14 Reduction of repetitions: ISI-1 best
Carpenter & Mueller (2013)	B/INT	Pronunciation	19 L1 English, L2 French	Reading and listening to 64 words (8 x 8 pron. rules)	Multiple choice pronunciation test	1 session	Blocked: 4 example words x rule (*bateau, fardeau, rameau*, etc.) Interleaved: 1 word x rule (*bateau, chacal, tandis*, etc.)	Immediate	Blocked practice better recognition of correct pronunciation
Collins et al. (1999)	PROG	General L2 skills	700 L1 French, L2 English Grade 6 (11-12)	Classroom instruction	Y/N vocab. recognition, MEQ (main emphasis listening), Written picture narrative	400 hrs	Massed: 5 months Massed plus: (+exposure) Distributed: 10 months	Immediate	Massed/massed plus > distributed
Collins & White (2011)	PROG	General L2 skills	230 L1 French, L2 English Grade 6 (11-12)	Classroom instruction	Vocabulary, Narrative writing, Listening skills, Oral interaction	300-400 h	Massed: 5 months Distributed: 10 months	Immediate	Massed > distributed in 6/20 comparisons Distributed > massed in 1/20
French et al. (2020)	PROG	Fluency, comprehensibility, Accentedness	81 L1 French, L2 English High School (Grade 10)	Classroom instruction	Picture-cue oral narrative task	One academic year	Intensive: 400 hrs in grade 6 Regular: 60 hrs	4 years	Intensive > Regular, Sig. more fluent and comprehensible speech four years after end of program No sig. diff. in accent
Kasprowicz et al. (2019)	LAG IND	Morphology (verb inflections number and tense)	113 L1 English, L2 French Grades 4-6 (8-11)	Digital app with mini-games focused on inflections	Sentence-picture matching, GJT Lang. analytic ability (LAA)	Long-spaced 3 of 60 mins Short-spaced 6 of 30 mins	ISI-3.5 ISI-7	RI-3.5 (for ISI-3.5) RI-7 (for ISI-7) RI-42	No sig. differences LAA predicted more learning in short lags
Koval (2019)	SP	Vocabulary (24 Finnish words in English sentences)	40 L1 English	Target words inserted in English sentences (reading)	Online processing, Form recognition, Form-meaning mapping	1 session (4 rep.) 2 hrs	Massed: subsequent rep. Spaced: 25 intervening sentences	48-72 hrs	Processing time: spaced > massed Vocab learning: spaced > massed
Koval (2022)	LAG/SP	Vocabulary (72 Finnish-English word pairs)	52 L1 English	L2-L1 retrieval of Finnish-English translation pairs	Form recognition L2-L1 translation Form-meaning matching	1 session (6 rep.) 3-4 hrs	Massed Short-spaced (17-38 trials) Long-spaced	Immediate Delayed (1-2 weeks)	Spaced > massed Long- > short-spaced delayed meaning posttest

							(71-119 trials)		
Küpper-Tetzel, Erdfelder & Dickhäuser (2014a)	LAG /SP	Vocabulary (26 German–English vocabulary pairs)	65 L1 German, L2 English Grade 6 (11-13)	Words presented via projector and read aloud (classroom)	L1-L2 cued-recall	2 sessions	Massed ISI-1 ISI-10	RI-7 RI-35	RI-7: ISI-1 > massed or ISI-10 RI-35: ISI-1 and ISI-10> massed
Küpper-Tetzel, Kapler & Wiseheart (2014b)	LAG	Vocabulary (28 word pairs)	210, L1 English University	Word pairs presented on a screen (individual)	Free recall	3 sessions	Contracting: ISI-5, ISI-1; Equal: ISI-3; Expanding: ISI-1, ISI-5 (days)	Immediate RI-1 RI-7 RI-35	Immediate: no sig. diff. RI-1 & RI-7: contracting > equal & expanding RI-35: equal & expanding > contracting.
Li & DeKeyser (2019)	LAG	Tonal word production (20 disyllabic words)	68 L1 English, L2 Mandarin University	1) presentation 2) declarative knowledge: meaning-spelling mapping 3) "procedural" oral prod	1) oral picture naming task 2) Written picture-naming task 3) Oral word-naming task	3 sessions	Cond A: ISI-1 Cond B: ISI-1 Cond C: ISI-7 Cond D: ISI-7	Cond A: RI-7 Cond B: RI-28 Cond C: RI-7 Cond D: RI-28	RI: strong effect retention of declarative knowledge (longer RI worse) RI-28: ISI-7 better than ISI-1 Procedural knowledge of new words: ISI important (ISI-1 better). ISI: 1-day better at pre-session performance btw sessions
Loftohali & Salehi (2017)	SP	Vocabulary (20 Eng.-Farsi word pairs)	28 L1 Farsi, L2 English, primary school	Classroom instruction	L1-L2 translation	40 mins 1 or 2 sessions	Massed: 1 session Spaced: 2 sessions	RI-7 RI-35	Spaced > massed, both RIs
Macis et al. (2021)	SP	Vocabulary (25 adjective + noun English collocations)	105 L1 Arabic L2 English (55 incidental; 50 deliberate)	Collocations inserted in a text (Exp. 1: incidental learning; Exp. 2: deliberate)	Cued form-recall test	5 sessions (15 min/session)	Massed: 5 words x 5 times per session Spaced: 25 words per session, ISI-7	RI-21 (for all words in spaced, but not for massed)	Incidental: massed > spaced > control Deliberate: spaced > massed > control
Miles (2014)	SP	Grammar (Frequency adverb-verb word order; the word *almost*)	32 L1 Korean, L2 English University	Classroom instruction	GJT (20 sentences), L1-L2 translation (19 sentences)	60 mins 1 session (massed) 3 sessions (spaced)	Massed Spaced: ISI-7, ISI-28 (average ISI-17)	Immediate RI-35 (5 weeks)	GJT: Immediate: no sig. diff.; Delayed: sig. advantage spaced Translation: no sig. diff. immediate or delayed
Nakata (2015)	SP/ LAG	Vocabulary (33 English-Japanese word pairs: 20 target +13 filler)	128 L1 Japanese, L2 English University	Paired-associate learning	Productive (L1-L2 translation) Receptive (L2-L1 translation)	1 session (4 rep.)	Massed Short-spaced (5-item) Medium-spaced (10-item) Long-spaced (30-item)	Immediate RI-7	Spaced > massed No diff. short, medium, long lags
Nakata & Elgort (2021)	SP	Vocabulary 48 pseudowords	66 L1 Japanese, L2 English University	Target pseudowords in Eng sentences (reading)	Meaning recall, Meaning-form matching, Semantic priming	1 session (96 mins) (3 rep.)	Massed Spaced: 47 items (25 min.)	Immediate Delayed: RI-2	Meaning recall and meaning-form matching: spaced > massed Semantic priming: no sig. diff.
Nakata & Suzuki (2019a)	SP	Vocabulary (48 English-Japanese pairs in 8 sets)	133 L1 Japanese, L2 English University	Paired-associate learning	L2-L1 translation	1 session, (4 rep.)	Massed by set Spaced	Immediate RI-7	Spaced > massed at both RIs and especially for unrelated sets
Nakata & Suzuki (2019b)	B/INT	Grammar (simple past, present perf., conditionals)	115 L1 Japanese, L2 English University	50 multiple choice fill-in-the-blank questions	GJT: 40 sentences	1 session Sets repeated 4 times	Blocked: 10A, 10B, 10C, 10D, 10E Interleaved (e.g.: A, B, C, D, E) Increasing: first 25 questions blocked; next 25 interleaved	Immediate RI-7	Immediate: no sig. diff. Delayed: Interleaved > Blocked Blocking more effective with low prior knowledge: interleaving more effective with high prior knowledge.

Nakata & Webb (2016)	LAG	Vocabulary (23 low-freq English words, incl. 3 fillers)	95 L1 Japanese, L2 English University	Paired-associate learning	Productive (L1-L2 translation) Receptive (L2-L1 translation)	1 session	Spaced: 3-item 19-item	Immediate RI-7	Long lags > short lags at RI-7
Rogers (2015)	LAG	Grammar (complex syntax / cleft sentences)	37 L1 Arabic? L2 English University	100 stimulus sentences, each followed by y/n comprehen. questions	20 GJT, English L2 grammar learning	5 sessions 15 mins each	ISI-2.25 ISI-7	RI-42	Sig. advantage for the long-spaced group
Rogers & Cheung (2020)	LAG	Vocabulary (20 adjectives)	52 L1 Cantonese, L2 English Grade 3	Teacher used pictures	Multiple choice meaning recognition	3 sessions 10 mins	ISI-1 ISI-8	RI-28	Spaced-short items learned sig. better than spaced-long
Rogers & Cheung (2021)	LAG	Vocabulary (20 words)	66 L1 Cantonese, L2 English Grade 4	Teacher used PPTs, pictures and crossword puzzles	Crossword puzzle production test (form recall)	3 sessions 10 mins	ISI-1 ISI-8	RI-28	No sig. differences between conditions
Schuetze (2015)	LAG	Vocabulary (39 English-German word pairs)	76 L1 English, L2 German University	Intentional learning of individual words presented on screen with audio	L1-L2 translation	Exp.1: 4 days Exp. 2: 5 days	Exp. 1: Uniform: ISI 3-4-3 Expanding: ISI 1-3-4 Exp. 2: Uniform: ISI 3-4-3-4 Expanding ISI 1-2-4-7	Immediate RI-28 RI-56	No sig. diff.
Schuetze & Weimer-Stuckmann (2011)	LAG	Vocabulary (40 words per textbook chapter)	117 L1 English, L2 German University	Typing and rehearsing words in online vocab. program	Online ViVo quiz + Print retention test	6-8 days	Uniform: ISI-2 Expanding: ISI 0-2-3	Short: RI 2-4 Long: diff. RIs (9-5 months)	Short RI: no sig. diff. Long RI: uniform > expanding
Serfaty & Serrano (2022)	LAG IND	Grammar (conditional and future perfect) Linguistic and learner-related difficulty	129 L1 Khmer L2 English, Secondary school (10-18)	Online flashcards, 16 sentences (8/structure)	Productive cued recall as in training with 16 novel sentences Proficiency test	2 sessions per structure	ISI-1 ISI-7	RI-7 RI-28	No lag effects ISI*proficiency: longer ISI better for higher prof. learners ISI*time-on-task: longer ISI better for faster learners
Serrano (2011)	PROG	General L2 skills (listening, grammar, reading, vocabulary, written and oral production)	152 L1 Spanish/ Catalan, L2 English University	Classroom instruction	Proficiency test, Written and oral narrative	80 hrs	Extensive: 2h/week, 7 months Intensive: 5h/day, 4 weeks	Immediate	Intermediate: intensive > extensive (few measures) Advanced.: no sig. diff.
Serrano & Huang (2018)	LAG	Vocabulary (36 words)	71 L1 Mandarin L2 English High school (15-16)	Repeated reading promoting incidental vocab learning	Meaning-recognition L2-L1 matching test	7 days	Intensive: ISI-1 Spaced: ISI-7	Immediate ISI/RI = 25%	Immediate: intensive > spaced Delayed: no diff.
Serrano & Huang (2021)	LAG	Vocabulary (36 words)	72 L1 Mandarin L2 English High school (15-16)	Repeated reading promoting intentional vocab learning	Meaning-recognition L2-L1 matching test	7 days	Intensive: ISI-1 Spaced: ISI-7	Immediate ISI/RI = 25%	Immediate: intensive > spaced Delayed: intensive > spaced
Serrano & Muñoz (2007)	PROG	General L2 skills (listening, grammar, reading)	76 L1 Spanish/ Catalan, L2 English University	Classroom instruction	Fill-in-blanks, Transformation, Written and oral narrative	65 hrs	Extensive: 2 hrs/week, 7 months Semi-intensive: 2-2.5 hrs/week 11-15 weeks Intensive: 5h/week, 5 weeks	Immediate	No btw-group differences Within group analyses: semi-int. & intensive > extensive
Serrano et al. (2015)	PROG	Formulaic sequences	124 L1 Spanish/ Catalan, L2 English University	Classroom instruction	Oral picture-based narrative	80 hrs	Extensive: 2h/week, 7 months Intensive: 5h/day, 4 weeks	Immediate	Beginner and intermediate: intensive > extensive Advanced: no sig. diff.
Snoder (2017)	LAG	Grammar	45 L1 Swedish, L2 English	Reading & writing tasks focused on	Productive recall (translation)	3 sessions	Spaced: ISI-6; ISI-9	Immediate	No sig. diff.

								RI-21	
		(verb-noun collocations)	Age: 16	target items			Intensive: ISI-1; ISI-2		
Suzuki (2017)	LAG	Morphology (*Supurango*, novel miniature language) 24 verbs	60 L1 Japanese, L2 English University	Computerized training tasks in 3 steps: 1) vocabulary learning 2) explicit grammatical explanation 3) oral practice	1) Vocab test 2) Rule application test 3) Present prog. Test	4 sessions	ISI-3.3 ISI-7	RI-7 RI-28	Short lag > long lag in accuracy at both RIs. No sig. diff. in speed
Suzuki (2018) (re-analysis of Suzuki, 2017)	LAG IND	Morphology Procedural knowledge	60 L1 Japanese, L2 English University	Computerized training tasks in 3 steps: 1) vocabulary learning 2) explicit grammatical explanation 3) oral practice	1) Rule application test 2) Present prog. test 3) Tower of London task	4 sessions	ISI-3.3 ISI-7	RI-7 RI-28	Procedural ability related to faster RT ISI-3.3 > ISI-7 in terms of CV
Suzuki (2019) (re-analysis of Suzuki, 2017)	LAG IND	Morphology Aptitude and WM	60 L1 Japanese, L2 English University	Computerized training tasks in 3 steps: 1) vocabulary learning 2) explicit grammatical explanation 3) oral practice	L2 ability: 1) Rule application test 2) Present prog. (PP) test LLAMA-F, Ospan task, LLAMA-B & LABJ-PA	4 sessions	ISI-3.3 ISI-7	RI-7 RI-28	7-day ISI: MMRA predicted 7 out of 9 rule-application test outcomes, and all PP test outcomes 3.3-day ISI: MMRA predicted immediate PP posttests outcomes No sig. effects of WM
Suzuki (2021b)	B/INT	Oral fluency	50 L1 Japanese, L2 English University	3 narrative oral tasks	2 diff. oral tasks	3 sessions	Blocked: AAA, BBB, CCC Interleaved: ABC, ABC, ABC	RI-1	Blocked practice AAA more sig. development of oral fluency: blocked sig. superior for 6 out of 9 measures on at least 2 of the 3 days.
Suzuki (2021a) (re-analysis Suzuki, 2021a)	B/INT IND	Fluency Individual differences	68 L1 Japanese, L2 English University	3 narrative oral tasks	2 diff. oral tasks Non-word repetition (NWR) Trail making (TMT) Llama_B	3 sessions	Blocked: AAA, BBB, CCC Interleaved: ABC, ABC, ABC	RI-1	Scores in NWR and Llama_B better predictors under blocked practice. Scores in TMT related to improvement under interleaved practice.
Suzuki & DeKeyser (2017a)	LAG	Morphology (Japanese - *te* present perf. verb form) 18 target verbs	51 L1 English, L2 Japanese University	1) Vocabulary learning 2) Explicit grammar 3) Sentence comprehension 4) Picture description 5) Narrative task	1) Rule application test 2) Sentence completion Accuracy and RT	2 sessions	ISI-1 ISI-7	RI-7 RI-28	Accuracy: no sig. diff. Speed: ISI-1 > ISI-7 long RI
Suzuki & DeKeyser (2017b) (re-analysis of 2017a)	LAG IND	Morphology Aptitude	40 L1 English, L2 Japanese University	Computerized training tasks (aural and visual)	L2 ability: 1) Rule application test: 18 action verbs 2) Sentence completion (using 18 target Vs) Aptitude: LLAMA_F, Ospan task	4 sessions	ISI-1 ISI-7	RI-7 RI-28	No correlation btw rule application & aptitude WM related to effectiveness of massed practice LAA related to effectiveness of distributed practice
Suzuki, Eguchi & de Jong (2022a) (re-analysis of Suzuki, 2021a)	B/INT	Fluency and use of constructions	50 L1 Japanese, L2 English University	3 narrative oral tasks	2 diff. oral tasks	3 sessions	Blocked: AAA, BBB, CCC Interleaved: ABC, ABC, ABC	RI-1	Blocked practice higher reuse of constructions.
Suzuki & Hanzawa (2022)	SP/ LAG	Oral fluency	70 L1 Japanese, L2 English University	One narrative picture prompted oral fluency task	3 diff. narrative picture prompted tasks	6 repetitions 1 or 2 sessions	Massed Short-spaced (45 min.) Long-spaced	Immediate RI-7	Training: no diff. at the end but massed more repetitions RI-7 Immediate post: massed positive for breakdown

Study	Type	Target	Participants	Training	Test	Duration	Conditions	Timing	Results
							(ISI-7)		fluency but negative for speed and repair. Delayed post: no diff.
Suzuki & Sunada (2020)	B/INT	Grammar (relative clause construction) Compreh. and production	129 L1 Japanese, L2 English University	64 items in 2 formats: a) Output-practice (oral picture description) b) Input-practice (aural comprehension picture selection)	16 items/test Accuracy and speed (same format as training materials)	1 session	Blocked: 16A, 16B, 16C, 16D Interleaved: C, B, D, A, C, etc Hybrid: 8A, 8B, 8C, 8D, then B, C, D, A, D, C...etc	Immediate RI-7	Immediate: Hybrid > blocked & interleaved (comprehension speed and production accuracy) RI-7: Hybrid > blocked & interleaved (comprehension speed) Input-practice > output-practice in interleaved & hybrid conditions
Suzuki, Yokosawa, & Aline (2022b)	B/INT IND	Grammar (English relative clauses, RC, 5 types: RC1, RC2, etc.)	60 L1 Japanese, L2 English University	Oral picture description 50 instances of 5 types of RC + feedback	Same as training with novel sentences (20 items) Listening span task	1 session	Blocked (RC1-RC1-RC1.... RC2-RC2, RC2...) Interleaved (RC1-RC2-RC3...)	Immediate RI-7	Immediate (accuracy and speed): interleaved > blocked Delayed (accuracy and speed): no diff. WM affects blocked but not interleaved practice.
White & Turner (2005)	PROG	Oral proficiency	152 L1 French, L2 English Grade 6 (11-12)	Classroom instruction	Self recording, Video story retell, Info-gap picture description	One academic year	Intensive: 5 months (300-400 h) Regular: 1 academic yr. (±60 h)	Immediate	Intensive > regular
Xu et al. (2014)	PROG	Oral proficiency	28 L1 English, L2 Mandarin High school (15-17)	Classroom instruction	SOPI (Simulated Oral Proficiency Interview)	85-88 hrs	Intensive: 4 weeks in summer Regular: 1 semester (22 weeks)	Immediate	Regular > intensive only in oral fluency
Yamagata et al. (2022)	SP	Verb-noun collocations	96 L1 Japanese, L2 English High school (15-16)	Presentation and production practice of target collocations in 7 stages	Collocation filling test Verb filling test	9 sessions, 3 sessions x 3 weeks (5-10 min./session)	Node massed Collocation massed Collocation spaced	Immediate RI-14	Spaced schedule better for learned and unlearned collocations

Abbreviations in alphabetical order:

B/INT: blocked/interleaved practice
IND: individual differences
ISI: inter-session interval
LAG: lag effects
PROG: program
RI: retention interval
SP: spacing effect