Applied Psycholinguistics **31** (2010), 635–650
doi:10.1017/S0142716410000172

Effects of distributed practice on the acquisition of second language English syntax

STEVE BIRD
United Arab Emirates University

Received: January 21, 2009 Accepted for publication: June 7, 2009

ADDRESS FOR CORRESPONDENCE
Steve Bird, Department of Linguistics, College of Humanities and Social Sciences, United Arab Emirates University, P.O. Box 17771, Al Ain, United Arab Emirates. E-mail: sbird@uaeu.ac.ae

ABSTRACT
A longitudinal study compared the effects of distributed and massed practice schedules on the learning of second language English syntax. Participants were taught distinctions in the tense and aspect systems of English at short and long practice intervals. They were then tested at short and long intervals. The results showed that distributed practice led to superior test scores on the long-term tests, indicating that the learning of second language syntax can benefit from distributed practice in a manner very similar to that reported for other skills and information types in the experimental psychology literature. Implications for intensive language-leaning programs and syllabus design in general are discussed.

The old adage "practice makes perfect" may be a simple truism for some aspects of foreign language learning, but it also raises many unanswered questions about how practice affects cognition and learning and how best to exploit practice time in study routines and syllabus designs. What kind of practice? How much practice? How frequently? The present study focused on the question of frequency, in particular, the effects of different practice schedules on grammar learning during a university-level English language proficiency course.

The focus on frequency of practice was motivated by studies in applied psychology demonstrating that practice schedules can be optimized for long-term learning (Cepeda, Vul, Rohrer, Wixted, & Pashler, 2008). Many studies have demonstrated that when there is a time gap between practice sessions, known as "distributed" practice, long-term memory for studied materials tends to be better than when practice is "clustered" or "massed"; that is, when little or no time intervenes between sessions (for reviews, see Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006; Dempster, 1988, 1996; Rohrer & Pashler, 2007). This beneficial effect on long-term memory has been referred to as a "distributed practice effect" or "spacing effect" (Rohrer & Pashler, 2007). The effect has been demonstrated for a range of information types and skills, for example, mathematics (Rohrer

© Cambridge University Press 2010 0142-7164/10 $15.00

https://doi.org/10.1017/S0142716410000172 Published online by Cambridge University Press

& Taylor, 2006), vocabulary learning (Bahrick, Bahrick, Bahrick, & Bahrick, 1993; Bahrick & Phelps, 1987; Bloom & Shuell, 1981; Dempster, 1987; Glenberg & Lehmann, 1980), reading comprehension (Reder & Anderson, 1982), learning concepts in biology (Reynolds & Glaser, 1964), and learning to associate names and faces (Landauer & Bjork, 1978). The effect has been found under intentional and incidental learning conditions (Challis, 1993), in recognition and recall tasks (Greene, 1989), and in implicit and explicit memory tasks (e.g., Greene, 1990). Most of these findings were obtained in relatively brief laboratory studies, and calls have been made to extend distributed practice research into more classroom-like conditions to explore potential benefits (Dempster, 1988). A number of such studies have been conducted and have generally reported spacing effects similar to results from laboratory studies (Rohrer & Pashler; 2007; Rohrer & Taylor, 2006; Seabrook, Brown, & Solity, 2005). For example, in Rohrer and Taylor (2006), college students were taught how to solve a mathematics problem and were then assigned to one of two practice schedules. In one experiment, participants either practiced 10 problems in a single massed practice session, or they practiced the 10 problems over two distributed practice sessions separated by 1 week. When both groups were tested 4 weeks after the final practice sessions, the distributed learning group's scores were twice as high as those for the massed practice group. Rohrer and Pashler (2007) argue that these results should apply to a wide array of information and skill learning contexts. They predict, for example, that foreign language-leaning courses that offer intensive instruction in a short period of time may be less beneficial than courses that offer more distributed learning over a longer period of time.

A number of studies of intensive and distributed learning in authentic foreign language education programs have been conducted, and they have tended to find the opposite result to that predicted by Rohrer and Pashler (2007; Collins, Halter, Lightbown, & Spada, 1999; Freed, Segalowitz, & Dewey, 2004; Lapkin, Hart, & Harley, 1998; Netten & Germain, 2004; Peters, 2000; Serrano & Munoz, 2007; Spada & Lightbown, 1989). For example, Serrano and Munoz (2007) compared three groups of English language learners all studying English as a foreign language for a total of 110 hr. The three study conditions were identical except in the way in which the total study time was distributed. The intensive group studied for 5 hr per day from Monday to Friday, the semi-intensive group studied 2 hr per day Monday to Thursday, and the extensive group (the distributed learning condition) studied 4 hr per week distributed over 2 days each week. It was found that the extensive group was the weakest on tests assessing listening, grammar, vocabulary, and reading at the end of the course. The fairly consistent conclusion across these studies has been that intensive, more massed instruction appears to be better than distributed practice when it comes to foreign language learning.

Why do we find this discrepancy between the many psychology studies showing consistently large benefits of distributed learning across a wide range of information types and skills, and those studies conducted in language-leaning classrooms showing no difference or even worse performance in distributed learning conditions? Serrano and Munoz (2007) speculate that foreign language learning might involve information and cognitive skills that are qualitatively different from the kinds of knowledge and skills tested in most distributed practice experiments, such

https://doi.org/10.1017/S0142716410000172 Published online by Cambridge University Press

as memorizing lists of words. However, as reviewed above, distributed learning effects have been found across an array of information and skill types, indicating that distributed learning effects are quite common in many domains. In addition, the two main theoretical accounts of distributed practice effects, which are known as "deficient processing" accounts and "encoding variability" accounts (Greene, 1992), point to cognitive mechanisms and environmental variables that should also affect language acquisition. Deficient processing accounts posit that massed repetitions lead to qualitatively poorer processing because learners tend to pay less attention to subsequent presentations of the new material, which is attributable to, for example, fatigue and boredom, or because of a false sense of confidence that the information has been fully learned before the session is complete (Hintzman, 1976). However, factors such as fatigue, boredom, or overconfidence seem likely to affect any learning, including second language learning. In contrast, encoding variability accounts argue that distributed practice usually involves some environmental change in each new practice context, and these changes lead to richer memory traces and better recall (Glenberg, 1979). It again seems very likely that language learners would also benefit from any such contextual enrichment of memory traces. There is no obvious reason why any of these mechanisms of cognition and learning should not affect foreign language learning in a manner similar to other kinds of learning. The discrepancy in the two sets of experimental results is puzzling for these reasons.

Are there any other factors that could account for the different results obtained in language-leaning studies? There appear to be several important experimental design variables worth considering. The main variable addressed in the present article is the relationship between the study time distribution and the time lag before the final test (Rohrer & Pashler, 2007). For clarity, this article adopts the terminology of applied psychology for two variables: first, the amount of time separating each practice session is referred to as the "intersession interval" (ISI); second, the time lag between the final study session and the posttest is referred to as the "retention interval" (RI). These two variables can interact in a number of important ways. For example, many learning studies have found that a short ISI is especially beneficial for short RI tests (Rohrer & Pashler, 2007). As an illustration of this phenomenon, consider the familiar study technique of cramming, a colloquial term for massed practice. In a massed practice study condition the ISI is zero or very small: there is little or no time lag between study episodes. Cramming is often done just before tests, when the RI is very short. Many educational practitioners and students will probably agree that cramming can be useful, especially for short-term retention of information. At the same time, cramming is typically regarded as a bad study strategy if one's aim is long-term retention for the learned information and skills. This intuition has been borne out by studies demonstrating equivalent and even superior performance for massed practice conditions with short RIs but not with long RIs (Bloom & Shuell, 1981; Rohrer & Taylor, 2006). For example, in Rohrer and Taylor's (2006) study summarized above, although the distributed learning group outperformed the massed practice group on a test 4 weeks after the final practice session, there was no difference between the groups' scores on an earlier test administered after a 1-week RI. In general, a short RI test tends to benefit disproportionately from a

https://doi.org/10.1017/S0142716410000172 Published online by Cambridge University Press

short ISI, and this principle is consistent with the finding that cramming seems to work better just before a test than it does for a test weeks or months later.

Returning to the second language acquisition research reviewed above, Serrano and Munoz (2007) administered their final proficiency test immediately at the end of the course of study, which is at a very short RI. If a short ISI tends to benefit a short RI, this predicts the outcome of Serrano and Munoz (2007) because the intensive program used the shortest ISI of all the experimental conditions, and the final tests came in the last few weeks of the course of study. The results of such studies have been taken as evidence of the relative merits of intensive language-learning programs compared to distributed learning programs, but these comparatively higher test scores may not extend to long-term RI tests. Rohrer and Pashler (2007, p. 185) argue from their own research that, "in the realm of life-long learning, immersion style foreign language courses are popular, yet their brevity, which prevents sufficient spacing, should produce deceptively high initial levels of learning followed by rapid forgetting." In the domain of foreign language learning it seems especially important to explore long-term learning gains because serious foreign language learners presumably aim to maintain their proficiency long after the exams have been completed. Although Serrano and Munoz (2007) did not administer a long RI test in their study, they acknowledge the importance of long RI testing when assessing distributed learning conditions.

Some language-learning studies have examined long-term effects of intensive programs (Lapkin et al., 1998; Lightbown & Spada, 1994), but the results are somewhat difficult to interpret for reasons related to the total amount of study time and the relationships between the ISI and RI. For example, Lightbown and Spada (1994) report research that compared two groups of teenage French-speaking learners of English (Grade 11 in the Canadian school system, ~15–16 years old), one group having taken intensive English when they were under 10 years of age (Grade 5, ~8–9 years old), the other group having had nonintensive English when they were in the same school grade. They found superior communicative abilities in the group that had taken the intensive English program. However, an important limitation of this study is that when participants were asked about their exposure to and use of English outside of school during the intervening years, the immersion group reported significantly more English-speaking friends, more part-time jobs in which they used English, and higher frequency of watching English language television and films. These differences between the two groups in exposure to and use of English during the intervening years makes it difficult to draw strong conclusions about the cognitive benefit of intensive versus distributed programs because the immersion group had significantly more exposure and practice than the other group prior to the final tests. There remains the possibility that when exposure and practice are held constant across groups, distributed learning effects might be observed.

In another study that tested the long-term effects of intensive and distributed language-leaning contexts, Lapkin et al. (1998) compared learning effects in Canadian elementary school aged learners who studied French for a total of 40 min per day over 10 months or 80 min per day over 8 months or in 0.5-day intensive instruction over 10 weeks. On posttests that evaluated listening, speaking, reading, and writing immediately at the end of the courses of study, there were no

https://doi.org/10.1017/S0142716410000172 Published online by Cambridge University Press

differences between the groups in listening and speaking, but the intensive condition showed significantly higher scores on reading and writing tests. In a follow-up test conducted at the beginning of the following school year, the intensive group maintained their superior test scores on the writing test but not the reading test, and all other tests showed nonsignificant differences. The authors conclude that the intensive program in general led to equivalent long-term benefits and, in the case of the writing test, even superior benefits. One of the concerns of this study was that an intensive schedule in the French program would create long time gaps between courses (∼8 months intervened between courses for the intensive group), and this long time gap might have led to significant forgetting compared to the schedule of 40 min per day over 10 months that had traditionally been used in the Canadian high school academic year. The results showed that the much longer RI for the intensive group had no negative effect on performance. This result seems to suggest that for long-term retention it makes little difference whether language-learning programs are intensive or distributed, a conclusion once again at odds with many psychology studies. More surprising is that, despite the much longer RI for the intensive group, their performance was in some cases better than the distributed group on the long-term tests. Why might this be the case?

As in the studies reviewed above, one potentially important difference is the relationship between ISI and RI. In their meta-analysis of 317 experiments from the psychology literature, Cepeda et al. (2006) conclude that the optimal separation of study episodes is not an absolute quantity; it instead depends on the RI. Put simply, if we want to know how frequently to study for optimized memory, we need first to determine for how long we want to remember the information. Once again, a short ISI tends to benefit a short RI, whereas a longer ISI benefits longer RIs (Rohrer & Pashler, 2007). In addition, however, the benefit of a longer ISI appears not to be monotonically related to RI; that is, simply increasing the time lag between practice sessions does not guarantee better long-term memory. For any given RI, there instead seems to be an optimal ISI. Cepeda et al. (in press) report an experiment in which participants studied Swahili–English word pairs. The ISI ranged from 5 min to 14 days, whereas the RI was held constant at 10 days. The ISI had a strong overall effect on test scores, but the optimal ISI for best test performance was 1 day. Increases in the ISI beyond 1 day resulted in increasingly poorer test performance for the 10-day RI. They conclude that at some point the time lag between study sessions begins to be too long for the given RI and performance begins to deteriorate. From their meta-analysis, Cepeda et al. (2006) conclude that optimal ISI should be 15% of the RI. Similarly, Rohrer and Pashler (2007) estimate an optimal ISI range of 10% to 30% of the RI.

This effect of the ISI/RI ratio may account for some of results reported in language-learning studies. Lapkin et al. (1998) did not control the ISI/RI ratio. For all three groups the ISI was 1 day (all groups studied French every day); the RI for the half-day group was 240 days (8 months), whereas the RI for the 40-min and 80-min groups was 60 days. All of these ISI/RI ratios were therefore less than 1% of the RI. If the ISI/RI ratio is critical to long-term test performance, this may help explain why, despite the much longer lag between study and follow-up test, the half-day intensive group scores were the same as the distributed schedule on most measures. What this study might have revealed is support for the idea that

https://doi.org/10.1017/S0142716410000172 Published online by Cambridge University Press

ISI or RI alone are not the critical factors in what is retained; instead, it is the ratio of ISI to RI that is crucial. Once we attend to the ratio we find more or less an equivalence in the two study conditions (ISI < 1% or RI), and this could account for the largely equivalent test scores on Lapkin et al.'s long-term test.

In summary, because no extant studies in the foreign language-leaning literature have controlled the ISI/RI ratio in the manner suggested by Rohrer and Pashler (2007), it is possible that language-leaning experiments have found results at odds with psychology studies not because language learning is fundamentally different from learning other intellectual skills but rather that the studies did not control the ISI/RI ratio to optimize distributed learning effects. Further investigation therefore seems warranted, and the present study attempted to do so by controlling the ISI/RI ratio.

Having argued for the potential importance of the ISI/RI ratio in accounting for the results of some language-leaning studies, it should be pointed out that other differences in design make studies such as Lapkin et al. (1998) difficult to compare directly with results of many psychology laboratory studies. One potentially crucial difference is the nature of the materials and tasks used for training and testing in the two kinds of study. Psychology studies typically use very constrained and well-defined tasks (e.g., solving a single kind of mathematical problem, vocabulary list learning, memory for faces), whereas language-learning studies often use more global tasks to measure language proficiency, for example, writing a magazine advertisement or a composition about the pros and cons of a debatable topic (see Lapkin et al., 1998). These latter kinds of task presumably require simultaneous use of syntax, vocabulary, and nonlinguistic cognitive skills: they require a general proficiency in using language, and they appear to be much more complex than those used in most psychology studies. There is evidence that increased task complexity can diminish distributed learning effects. Donovan and Radosevich (1999) compared distributed learning effects in tasks designed to be relatively simple or complex conceptually. They found that the magnitude of distributed learning gains varied according to the given task's conceptual difficulty. For example, distributed learning showed greater benefit for word-list learning (rated by the authors as average conceptual difficulty) than for puzzle solving (rated by the authors as high conceptual difficulty). Thus, an important caveat for any distributed learning study seems to be that not all tasks are necessarily affected equally by manipulations of study time distribution. Global measures of proficiency may involve a level of complexity that eliminates distributed learning effects. Conversely, foreign language vocabulary learning has been shown to be beneficially affected by distributed practice (Bahrick et al., 1993; Bloom & Shuell, 1981), and this may reflect the relatively uncomplicated nature of vocabulary learning (e.g., word lists with native language translations; Cepeda et al., 2007). The present research put aside issues related to global level proficiency and instead concentrated on a more narrowed investigation to explore whether there are any other kinds of linguistic information and skill that can be beneficially affected by distributed learning. Here we focused on grammar. The research question was as follows: Can distributed practice affect the ability of learners of English to correctly distinguish the use of simple past, present perfect, and past perfect verb forms on short- and long-term proficiency tests?

https://doi.org/10.1017/S0142716410000172 Published online by Cambridge University Press

Based only on the evidence from the psychology literature, the answer should be yes. We should see distributed practice benefits, at least at long RIs, where the ISI is between 10% and 30% of the RI. However, this prediction is moderated by the fact that the task involves a fair amount of conceptual difficulty, a factor that has been shown to reduce the magnitude of distributed learning effects. Furthermore, given the mixed findings for distributed learning effects in language-learning contexts, there was good reason to be uncertain whether distributed learning would yield effects similar to those reported in the psychology literature. It should be made clear from the outset that the design of this experiment precludes easy comparisons of its results with those of studies measuring global proficiency because the information and tasks used here tap only a very limited part of the information and cognitive mechanisms involved in the full complexity of language learning. The aim in the present study was conservatively directed only at exploring whether language-learning subskills and knowledge types that are somewhat more abstract and complex than vocabulary lists can be affected by distributed practice.

METHOD

Participants and courses

The participants were 38 learners of English enrolled in a university degree program in economics and business. These 38 came from two groups of 19 students. The participants were all first-year students ranging in age from 19 to 23 years. Their native language was Malay. Their English proficiency level at entry to the course was 5.8 to 6.0 on the International Standardized Test of English Language Proficiency, an internationally recognized test of English proficiency. These scores indicate roughly an intermediate proficiency level relative to a native speaker. In the participants' language community English is widely used in everyday communication, but the kind of English tends not to make consistent and reliable use of English verb morphology for past tense (e.g., I go/went) or perfective aspect in the past (e.g., I have gone/had gone). These features of their English are often a focus of remedial work in their compulsory English proficiency courses because, although the learners have typically been taught these forms in school, many have never mastered the formal and conceptual distinctions.

The courses were both a total of 28 hr of class time divided into 3 hr per week (one 2-hr session plus one 1-hr session) over a 14-week semester. One instructor taught both classes.

Materials

The study used isolated, form-focused materials (Spada, 1997). That is, the grammatical forms were presented and practiced out of context of other topics and tasks used in the overall course. This was done in part to give experimental control over what all participants were paying attention to and for how long in the study sessions. Although these materials and the form-focused approach are not the only ones that could have been used, much research has demonstrated their benefits, so

https://doi.org/10.1017/S0142716410000172 Published online by Cambridge University Press

Table 1. *Ratios of intersession intervals to retention intervals*

	7-Day RI	60-Day RI
3-day ISI	42%	5%
14-day ISI	200%	23%

Note: RI, retention interval; ISI, intersession interval.

they were also selected for their educational merit (for reviews, see Bardovi-Harlig, 2000).

The learning materials were two sets of grammar worksheets (see Appendix A for a sample sheet). The first set included 160 sentences made up of 80 simple past (e.g., *I saw that movie with my brother last week*) and 80 present perfect sentences (*I have seen that movie many times*). The second set included 160 sentences made up of 80 present perfect sentences and 80 past perfect sentences (e.g., *I hadn't seen that movie before she recommended it*). Each of the 320 sentences was unique. The sentences were then divided into 16 sets of 20 sentences: 8 sets of simple past/present perfect (SP/PP) sentences and 8 sets of present perfect/past perfect (PP/PP) sentences. In each set of 20, 15 sentences were randomly selected to be erroneous sentences, and the main verb phrase was altered to make its tense or aspect grammatically incorrect. Thus, in each set of 20 sentences, 15 contained a tense or aspect error (e.g., *I have seen* that movie with my brother last week*) and 5 were correct. The 16 sets of sentences (8 SP/PP sets, 8 PP/PP sets) were then printed on separate A4 sheets of paper, and each sheet was used either as a worksheet ($N = 10$) or as a test sheet ($N = 6$). For each worksheet there was an overhead transparency showing the sentences and indicating which were correct or incorrect, with the errors underlined and the corrections for the erroneous sentence written below each sentence. These transparencies were the basis for the instructor's explanation to the participants after they had completed the task.

Design

Participants were randomly assigned as intact groups to experimental conditions. Both groups underwent the same TASK, ISI and RI conditions, so these were the within-subjects factors. TASK included two levels: SP/PP and PP/PP. ISI included two levels: a 3-day ISI and a 14-day ISI. RI included two levels: a 7-day RI and a 60-day RI. The ISI and RI time lags were chosen in order to manipulate the ISI/RI ratio within the constraints of a 14-week semester and the course syllabus's other task and time requirements. The combinations of ISI and RI created four ISI/RI ratios, which are shown in Table 1.

The values represent ISI as a percentage of RI. Note that the only ISI that falls within Rohrer and Pashler's (2007) optimal range of 10% to 30% of the RI is the 14-day ISI/60-day RI pairing. This was a critical experimental condition because it tested the effect of an optimized ISI for a long lag RI. The 7-day RI condition

https://doi.org/10.1017/S0142716410000172 Published online by Cambridge University Press

represents the type of testing condition reported in numerous language-learning studies, which is a final examination administered shortly after the end of the course. In this condition, as summarized above, most language-learning studies have found that the shorter ISI (as in intensive courses) results in higher test scores than distributed conditions (i.e., conditions with longer ISIs). The 3-day ISI was well out of the optimal range for the 60-day RI, and it tested the effects of intensive practice on long-term learning. The 14-day ISI/7-day RI condition also tested the effect of a longer ISI with a short RI, where the ISI is again well out of the optimal range. The 3-day ISI/7-day RI ratio is just outside the optimal range, and it tested the effect of a short ISI on a short RI, which are the conditions used in many intensive language-learning studies when comparing intensive and distributed learning schedules.

One group of participants was assigned to a task condition in which SP/PP practice was on a 3-day ISI schedule, whereas the PP/PP was on the 14-day ISI schedule. The reverse task–schedule pairing was used with the second group of participants to allow comparison of the tasks on different ISI schedule. Thus, the TASK-ISI pairing was a between-subjects variable.

Procedure

Both groups were given the pretest on the first day of the course. The test comprised one of each of the worksheets from each set; hence, each participant was given a sheet of 20 SP/PP sentences and a sheet of 20 PP/PP sentences. The instructions were written at the top of the pages: participants were asked to read each sentence and judge whether the sentence was grammatically correct (for each set of 20 sentences, 15 were correct and 5 were incorrect; participants were not told this). If a sentence was judged correct, participants were to place a tick next to it and go on to the next sentence. If it was not correct, they were to rewrite the sentence in the space below to make it grammatically correct. Participants were given 1 mark for correctly identifying a correct/incorrect sentence and another mark for making the appropriate correction. Thus, their score on the test was out of 35 marks: 5 marks for the correct sentences and 30 marks for the 15 incorrect sentences. The participants were given 1 hr to complete the test.

In the study phase, both groups received both sets of task sheets over the course of the semester. As discussed above, the only difference in procedure between the groups was the practice schedule assigned to the sets of work sheets. Group 1 received the SP/PP worksheets every 3 days and the PP/PP worksheets every 14 days. Group 2 received the reverse schedule for the same tasks (again, to counterbalance the tasks and schedules). The study sessions occurred within the 14-week semester. There were a total of five study sessions for each of the two verb form conditions (SP/PP and PP/PP). Each practice session lasted 1 hr. In each session, participants were each given a worksheet and told that they had 30 min to complete it. To maintain focus on the verb forms of interest, participants were reminded that the only error in any of sentences would be in the verb tense or aspect, and that no sentence should be judged as incorrect for any other reason (see the instructions for the worksheet in Appendix A). After 30 min the transparency was presented on the overhead projector, and participants were given the correct

https://doi.org/10.1017/S0142716410000172 Published online by Cambridge University Press

answers as well as brief explanations of why each verb phrase in each sentence was correct or incorrect and how to form the correct sentence. The oral explanations given for each sentence were identical for each group (e.g., *I have seen that movie with my brother last week* is wrong because the simple past is used when we know when the completed action happened) and were related only to the verb tense and aspect morphology of concern in the experiment. To control the amount of practice time and restrict it to the 1-hr classroom sessions, no explicit instruction on these grammatical forms was given during the remaining time in class (during the 2-hr sessions). The remaining time was instead given to listening, speaking, and writing tasks unrelated to the grammar exercises. At no time were the participants asked to study these sheets outside class, nor were they told that there would be posttests at the end of the course. Participants were also told in the first week of the course that the grammar exercises were supplementary and would not be included on the final examination. Finally, all homework assignments addressed topics unrelated to the grammar tasks (see post hoc interviews below for a discussion of extracurricular practice).

Two posttests each comprising 20 SP/PP and 20 PP/PP sentences were administered to each group. For both groups, the first test was given at a 7-day RI after the group's final study session. The second test was given at a 60-day RI after each respective group's final study session. Both were surprise tests for all participants. The 60-day RI test was administered after the participants had had an intersemester break between courses and thus was especially surprising because the course had been completed and the participants were enrolled in a new course in a new semester and studying new topics and tasks unrelated to the experiment. The researcher was the instructor for the second semester courses, and the syllabus was designed to avoid extra practice of the experimental tasks.

Following the 7-day RI test administered at the end of the course, the researcher asked each classroom group whether they had noticed the different study schedules for the different tasks over the semester and whether they had spent time practicing the tasks outside the classroom. This was done to check whether any group had had extra practice time that might have affected the final results. (No extra practice time was reported; see the Results Section.)

RESULTS

The results of the pretest and posttests are provided in Table 2 and Figure 1. As indicated in the figure, 20 of the 35 marks were given for the correct–incorrect part of the test; hence, chance performance on the test was 10. An initial analysis of variance (ANOVA) showed no significant differences between the groups on the pretests ($p > .05$); hence, the pretest results in Table 1 and Figure 1 are presented as mean scores collapsed across groups (a total of 38 participants). Looking just within each condition (i.e., each line on the graph), a repeated-measures ANOVA with least significant difference pairwise comparisons showed significant differences in test scores within all TASK-ISI conditions: SP/PP 3-day ISI, $F\,(1, 18) = 25.91, p < .001, \eta_p^2 = 0.590$; SP/PP 14-day ISI, $F\,(1, 18) = 85.31$, $p < .001, \eta_p^2 = 0.826$; PP/PP 3-day ISI, $F\,(1, 18) = 46.64, p < .001, \eta_p^2 = 0.722$;

https://doi.org/10.1017/S0142716410000172 Published online by Cambridge University Press

Table 2. *Scores by retention intervals for 3- and 14-day intersession intervals*

	Pretest (*SD*)[a]	7-Day RI (*SD*)	60-Day RI (*SD*)
SP/PP			
3-day ISI	18.82 (2.74)	29.10 (2.21)	22.21 (2.02)
14-day ISI	18.82 (2.74)	28.33 (4.51)	26.89 (1.38)
PP/PP			
3-day ISI	13.22 (1.88)	24.12 (5.20)	17.35 (2.38)
14-day ISI	13.22 (1.88)	22.23 (4.42)	21.20 (3.33)

Note: RI, retention interval; ISI, intersession interval; SP/PP, simple present/present perfect task; PP/PP, present perfect/past perfect task.
[a]Means are collapsed across both groups ($N = 38$) because there were no significant differences between groups on pretest scores.

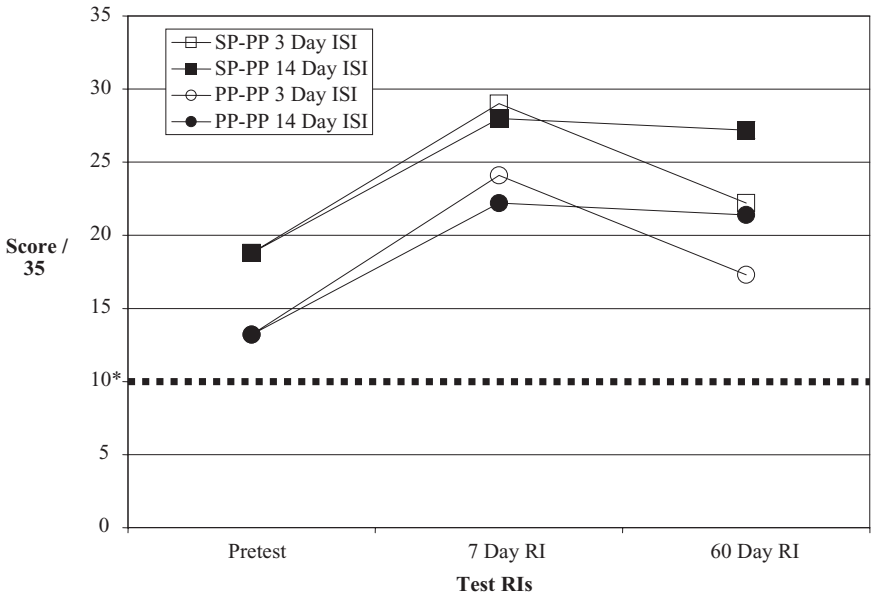


Figure 1. Test scores by retention intervals (RIs) for 3- and 14-day intersession intervals (ISIs).
*The dotted line indicates chance-level performance.

PP/PP 14-day ISI, F (1, 18) = 271.83 $p < .001$, $\eta_p^2 = 0.938$. Least significant difference comparisons showed that all 7-day posttest scores were significantly higher than their respective pretests ($p < .05$). Comparison of the 7-day and 60-day test results revealed the following: first, in the 3-day ISI condition, scores on the 60-day tests were significantly lower than for their respective 7-day test ($p < .05$); second, in the 14-day ISI condition there were no significant differences between the 7-day and 60-day posttests.

https://doi.org/10.1017/S0142716410000172 Published online by Cambridge University Press

Comparing between the two groups' TASK-ISI conditions, planned comparisons revealed the following: at the 7-day RI test, the 3- and 14-day ISI groups showed no significant score differences for either task ($p > .05$); for the 60-day RI test there was a significant difference between the 3- and 14-day ISIs for the SP/PP task, $t(18) = 5.87$, $p < .001$, $d = 2.70$, and for the PP/PP task, $t(18) = 2.81$, $p < .05$, $d = 1.33$. As shown in Figure 1, for both task conditions the 3-day ISI scores were significantly lower than the 14-day ISI scores on the 60-day posttest. Comparing across the linguistic tasks, note in Figure 1 that the same group of participants that sustained gains at the 60-day interval with a 14-day ISI were also the group that for the other linguistic task did not sustain the gains with a 3-day ISI. This pattern suggests that the crucial variable was the ISI for both groups of participants.

As described earlier, following the 7-day RI test at the end of the course, each classroom group was asked orally by the researcher whether they had noticed the different study schedules for the different tasks over the semester. None had. They were also asked whether they had spent time practicing the tasks outside the classroom. Three participants from one group and four from another reported having briefly consulted the sheets when preparing written work for other courses, but no participant reported having spent more that a few minutes studying outside of the class time. All participants reported that because the materials did not count toward their final grades, they had seen no need to practice further. This information was taken as confirmation that the groups had had equivalent practice time over the semester and that the practice had overwhelmingly been confined to the classroom sessions.

DISCUSSION

The critical comparisons in this study were between two levels of ISI at two different RIs. The results showed that for both ISI conditions there were significant test performance improvements at the 7-day RI relative to their respective pretest score. However, the two ISI conditions scores were equivalent, indicating that studying at a 3-day or 14-day ISI made no difference on the 7-day posttest performance. Therefore, no distributed learning effect was found on the short lag posttest. This result is consistent with studies showing no difference between intensive and distributed language-learning programs on final tests at the end of the course (e.g., Lapkin et al., 1998). However, these gains were only sustained by the 14-day distributed learning condition on the 60-day posttest. For the 3-day ISI condition on both language tasks, scores were significantly lower than the respective 7-day test score. That is, after 60 days a significant amount of what had been learned during training was forgotten in the 3-day ISI condition but not in the 14-day ISI condition. Moreover, when comparing between conditions at 60 days, the 14-day ISI condition scores were significantly higher than the 3-day ISI scores, whereas they were equivalent at 7 days. This again demonstrates a significant amount of forgetting in the 3-day ISI condition relative to the 14-day ISI. We can conclude that the distributed learning conditions yielded better long-term retention of what had been learned during the study phase. The implications of these results are discussed in turn below.

https://doi.org/10.1017/S0142716410000172 Published online by Cambridge University Press

These results are consistent with the large body of evidence in the psychology literature reporting that distributed practice can have a beneficial effect on long-term learning across a variety of information and skill types, provided that the ISI and RI are set to a ratio where ISI is 10% to 30% of the RI (Cepeda et al., 2006; Rohrer & Pashler, 2007). They are also consistent with the more fine-grained observation that when RI is short, ISI can have little or no effect on test performance: the ISI conditions did not result in differences in test scores at a shorter 7-day RI. Only when the ISI was 23% of the RI (14-day ISI/60-day RI) was a distributed learning effect observed. Thus, the present study demonstrates that when ISI and RI ratio are controlled appropriately, some aspects of foreign language grammar learning show a distributed learning effect similar to that observed in other learning domains (Dempster, 1987; Glenberg & Lehmann, 1980; Rohrer & Pashler, 2007).

Turning to the language-learning literature, these results are both consistent and inconsistent with distributed learning studies. The nonsignificant difference between test scores at a 7-day RI is consistent with those studies showing no benefit of longer ISIs on language learners when final tests are administered immediately at the end of a course of study (Collins et al., 1999; Freed et al., 2004; Lapkin et al., 1998; Netten & Germain, 2004; Peters, 2000; Serrano & Munoz, 2007; Spada & Lightbown, 1989). The conclusion then is that with respect to end of course tests, the short RI appears not to favor distributed learning conditions. However, the effect of ISI at the longer 60-day RI is inconsistent with the broader conclusion made in some studies that distributed learning does not seem to affect foreign language learning generally (e.g., Serrano & Munoz, 2007). What the results here indicate is that ISI can affect core aspects of foreign language learning (grammar) in a beneficial manner similar to that observed for other intellectual skills and information types, particularly in terms of sustaining gains over longer RIs. As has been demonstrated for foreign language vocabulary learning (e.g., Bahrick et al., 1993), in the context of distributed learning effects there therefore appears to be no reason to posit (as in Serrano & Munoz, 2007) a dissociation between the cognitive processes involved in foreign language grammar learning and the cognitive processes involved in other kinds of information and skill learning that have been shown to be affected by distributed practice.

However, as discussed in the introductory section, we must be careful about generalizing from the present results to global measures of proficiency. Some authors have reported that distributed practice effects depend on the nature of the task, particularly the level of complexity involved (Donovan & Radosevich, 1999). One concern from the outset of this study was the extent to which the complexity of the grammatical distinctions might counteract the benefit of distributed practice. Results showed that the ability to correctly detect errors and generate correct syntactic form distinctions for the simple present, simple past, present perfect, and past perfect improved and was sustained over a long period of time in distributed practice conditions. These improvements admittedly only pertain to subsets of the knowledge and cognitive skills required for tests measuring overall proficiency (e.g., overall communicative ability; Lapkin et al., 1998). As suggested by results of some language-learning studies, there remains the possibility that at the global proficiency level, language learning involves a degree of complexity that eliminates the benefit of distributed learning regardless of how the ISI/RI ratio is set (Collins

https://doi.org/10.1017/S0142716410000172 Published online by Cambridge University Press

et al., 1999; Freed et al., 2004; Lapkin et al., 1998; Netten & Germain, 2004; Peters, 2000; Serrano & Munoz, 2007; Spada & Lightbown, 1989). Speculation as to why complexity of global proficiency tests might have this effect is beyond the scope of the present article. Future research can gain clarity in this area by controlling the kind and complexity of linguistic tasks as well as the ISI/RI ratio, as in the present study.

One related limitation of the present study that should be pointed out is that the tests measured the ability to detect and correct verb morphology in a context that enforced conscious attention to the verb forms. Full language proficiency, of course, depends on the ability to produce grammatically well-formed sentences at a discourse level spontaneously and correctly during real communication with little or no conscious attention to verb morphology. The present results do not warrant a conclusion that the grammatical distinctions have been "learned" in this more global sense of language learning. As above, it is possible that, for example, developing an automatized ability to effortlessly produce grammatically well-formed sentences in real communication happens to be an ability that can benefit from intensive instruction. This, of course, is speculation and needs investigation.

In terms of practical implications, the present results indicate that language learners can benefit in the longer term from form-focused instruction with a distributed schedule of practice. Although numerous studies have demonstrated merits of more time-intensive programs (although many questions remain concerning these studies because ISI–RI were not controlled), we can at a more fine-grained level see some practical applications of the present results. What we can say is that within any language-learning program, intensive or distributed, revision of grammatical information and skills at carefully controlled intervals seems to help learners sustain proficiency gains over a longer period of time after study. The implication for language syllabus design is that the benefits of repeated practice of earlier learned material (reviews, quizzes, etc.) may be enhanced by controlling the period of time intervening between these sessions. As in other knowledge and skill-learning domains, frequency of practice therefore seems to be a factor worth attending to in foreign language learning. A challenge now is to determine the parameters of distributed learning effects within authentic foreign language-learning environments.

APPENDIX A

Sample worksheet (simple past–present perfect)

For each sentence below, decide whether it is grammatically correct. If it is, place a tick next to the sentence. If it is wrong, place an X and rewrite the sentence so it is grammatically correct.

1. Yamin got an A grade in this course last semester.
2. When have you arrived?
3. We have left Singapore in 1997.
4. Tony has quit his job last May.
5. Thomas Edison has invented the light bulb.

https://doi.org/10.1017/S0142716410000172 Published online by Cambridge University Press

6. Sheena has lived in France from 1998 to 2000.
7. My mother has made a great cake last weekend.
8. Science made many important advances in 20th century.
9. Narissa missed three classes so far this semester.
10. My parents have lived in their present home since 1976.
11. My father has visited Italy just before he died.
12. Last Christmas I have bought a car for my daughter.
13. I worked at this university for 2 years already.
14. I've never seen a camel in my life.
15. I always loved chocolate since I was a little boy.
16. How long did you sit in this classroom so far this morning?
17. Have you make any money on your investments so far this year?
18. Grant has just finished his breakfast.
19. David has left for lunch at 11:30.
20. Albert Einstein has been a great mathematician.

REFERENCES

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

Bardovi-Harlig, K. (2000). *Tense and aspect in second language acquisition: Form, meaning, and use*. Oxford: Blackwell.

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

Cepeda, N. J., Mozer, M. C., Coburn, N., Rohrer, D., Wixted, J. T., & Pashler, H. (in press). Optimizing distributed practice: Theoretical analysis and practical implications. *Experimental Psychology*.

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

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, 11*, 1095–1102.

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

Collins, L., Halter, R. H., Lightbown, P. M., & Spada, N. (1999). Time and the distribution of time in L2 instruction. *TESOL Quarterly, 33*, 655–668.

Dempster, F. N. (1987). Effects of variable encoding and spaced presentations on vocabulary learning. *Journal of Educational Psychology, 79*, 162–170.

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. (1996). Distributing and managing the conditions of encoding practice. In E. L. Bjork & R. A. Bjork (Eds.), *Memory* (pp. 318–339). New York: Academic Press.

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.

Freed, B. F., Segalowitz, N., & Dewey, D. P. (2004). Context of learning and second language fluency in French: Comparing regular classroom, study abroad, and intensive domestic immersion programs. *Studies in Second Language Acquisition, 26*, 275–301.

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., & Lehmann, T. S. (1980). Spacing repetitions over 1 week. *Memory & Cognition, 8*, 528–538.

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

Greene, R. L. (1990). Spacing effects on implicit memory tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 1004–1011.

https://doi.org/10.1017/S0142716410000172 Published online by Cambridge University Press

Greene, R. L. (1992). *Human memory: Paradigms and paradoxes*. Hillsdale, NJ: Erlbaum.
Hintzman, D. L. (1976). Repetition and memory. In G. Bower (Ed.), *The psychology of learning and memory* (pp. 47–91). New York: Academic Press.
Landauer, T. K., & Bjork, R. A. (1978). Optimum rehearsal patterns and name learning. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory* (pp. 625–632). London: Academic Press.
Lapkin, S., Hart, D., & Harley, B. (1998). Case study of compact core French models: Attitudes and achievement. In S. Lapkin (Ed.), *French second language education in Canada: Empirical studies* (pp. 3–31). Toronto: University of Toronto Press.
Lightbown, P. M., & Spada, N. (1994). An innovative program for primary ESL students in Quebec. *TESOL Quarterly, 28*, 563–579.
Netten, J., & Germain, C. (2004). Theoretical and research foundations of intensive French. *Canadian Modern Language Review, 60*, 275–294.
Peters, M. (2000). Le bain linguistique: Une innovation pour le programme de Français de base. *Apprentissage et Socialisation, 20*, 71–86.
Reder, L. M., & Anderson, J. R. (1982). Effects of spacing and embellishment on memory for the main points of a text. *Memory & Cognition, 10*, 97–102.
Reynolds, J. H., & Glaser, R. (1964). Effects of repetition and spaced review upon retention of a complex learning task. *Journal of Educational Psychology, 55*, 297–308.
Rohrer, D., & Pashler, H. (2007). Increasing retention without increasing study time. *Current Directions in Psychological Science, 16*, 183–186.
Rohrer, D., & Taylor, K. (2006). The effects of overlearning and distributed practice on the retention of mathematics knowledge. *Applied Cognitive Psychology, 20*, 1209–1224.
Seabrook, R., Brown, G. D. A., & Solity, J. E. (2005). Distributed and massed practice: From laboratory to classroom. *Applied Cognitive Psychology, 19*, 107–122.
Serrano, R., & Munoz, C. (2007). Same hours, different time distribution: Any difference in EFL? *System, 35*, 305–321.
Spada, N. (1997). Form-focussed instruction and second language acquisition: A review of classroom and laboratory research. *Language Teaching, 30*, 73–87.
Spada, N., & Lightbown, P. M. (1989). Intensive ESL programmes in Quebec primary schools. *TESL Canada Journal, 7*, 11–32.

https://doi.org/10.1017/S0142716410000172 Published online by Cambridge University Press