Effects of distributed practice on the proceduralization of morphology.

Yuichi Suzuki & Robert DeKeyser
Kanagawa University & University of Maryland

Abstract

The present study aimed to examine whether distributed practice works better than massed practice for proceduralization of grammatical knowledge. Learners of Japanese as a second language were trained on an element of Japanese morphology under either massed or distributed practice conditions. Results showed that massed practice led to accurate utterances to the same extent as distributed practice. It was also shown that massed practice may lead to more rapid utterances than distributed practice. A number of potential factors that mediate the effects of distributed practice are discussed.

I Introduction

One of the ultimate goals of second language (L2) learning is to attain fast, accurate, spontaneous, and effortless use of knowledge, i.e., automatic knowledge (DeKeyser, 2007). The present study aimed to explore what the optimal conditions are for proceduralization or automatization of grammatical knowledge to take place in the foreign language classroom, in the framework of Skill Acquisition Theory (SAT) (Anderson et al., 2004; DeKeyser, 2015). SAT stipulates that second language learners can first learn declarative knowledge (i.e., knowledge about the grammatical rules), and then engage in deliberate practice and develop procedural knowledge (i.e., knowledge about how to use the rules) that allows them to use a second language faster and more effortlessly. This procedural knowledge can be automatized with further extensive practice. These stages cannot be skipped or reversed; automatization requires procedural knowledge, and procedural knowledge requires declarative knowledge.

A practically important question for proceduralization/automatization of L2 skills is to investigate how L2 learners' practice schedules should be arranged. A large body of literature in cognitive psychology has explored whether long time intervals between training sessions (i.e., distributed practice) lead to a better retention of a variety of skills than short intervals (i.e., massed practice) (e.g., Carpenter et al. 2012). There is, however, strikingly little research on the effects of distributed practice in SLA, especially in areas of grammar learning (Bird, 2010). The present study aimed at investigating whether distributed practice enhances the proceduralization of morphological knowledge in Japanese as a second language. From the foreign language educator's vantage point, if two sessions are available to have students practice a morphological pattern, then is it better to schedule two sessions in close succession (e.g., within the same week) or to leave a longer time interval between them (e.g., next week)? In what follows, we will first review research on distributed practice in cognitive psychology. Next, we will delve into research on distributed learning of L2 skills and narrow down potential factors that influence the effectiveness of distributed practice.

II Background

1 Cognitive and educational psychology

In the cognitive psychology literature, it is widely accepted that learned knowledge is retained better when the practice is distributed rather than massed—this is referred to as the distributed practice effect. Much empirical research has supported this finding, both in laboratory (see Carpenter, et al., 2012; Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006; Toppino & Gerbier, 2014; for review) and classroom settings (Carpenter, Pashler, & Cepeda, 2009; Seabrook, Brown, & Solity, 2005; Sobel, Cepeda, & Kapler, 2011). For instance, Sobel et al. (2011) examined whether English-speaking children benefit from spacing of one-week interval (distributed) than 1-minute interval (massed) in learning unfamiliar English words. Results from vocabulary recall tests showed that they remembered the learned words better in the distributed learning condition than in the massed learning condition.

Of particular interest to the present study, the psychology literature suggests that the optimal interval between the practice sessions can be determined by the ratio of the space between them—the Inter-Session Interval (ISI)—and the time lag between the end of the practice session and the time of testing—Retention Interval (RI) (Cepeda et al., 2009; Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006). Rohrer and Pashler (2007) reported an empirical study that aimed at identifying optimal spacing. They examined the retention of trivial facts by manipulating the ISIs (up to 15 weeks) and RIs (as long as 50 weeks). Results suggested that the optimal ISI ranges from 10% to 30% of the RI; this ratio will be utilized in the current research design.

2 Second language acquisition research

The effects of distributed learning have also been evaluated extensively in second language programs in Canada and foreign language programs in Spain (see Serrano, 2012 for review). A few recent studies carefully manipulated the distribution of instruction hours across L2 learners' groups (Collins & White, 2011; Serrano, 2011; Serrano & Muñoz, 2007).[1] Results from these studies did *not* find any advantages for the distributed practice over the intensive practice, which contradicts the findings in the cognitive psychology experiments reviewed earlier. One potential explanation for the different findings in the two research paradigms is what type of target skills are learned. The program-level L2 studies primarily focus on the acquisition of more global English skills such as listening and reading performance, which involve complex processes (e.g., integration of lexicon and grammar for reading comprehension), whereas cognitive psychology studies often deal with simple cognitive tasks such as verbal recall.

While the program-level research focused on global L2 skills, distributed practice effects have also been examined with focus on L2 acquisition of discrete target items like vocabulary and grammar. Experimental studies have focused mostly on vocabulary learning, i.e., memorization of paired associates (Bloom & Shuell, 1981; Karpicke & Bauernschmidt, 2011; Nakata, 2015; Pavlik & Anderson, 2005; Rohrer & Pashler, 2007; Schuetze, 2015), and results demonstrated that distributed learning led to a better performance than massed practice.

Little research has been conducted on L2 grammar learning. An exception is Bird's (2010) study on the acquisition of English morphosyntax, which also confirmed the advantage in distributed practice over massed practice. Bird (2010) compared the acquisition of tense/aspect of the verb forms (i.e., simple past/present perfect and present/past perfect) under distributed and massed conditions. In both practice and tests, participants were given worksheets and told to indicate whether tense/aspect of the verb forms were correct or not. They were required to check whether the form of verbs was used correctly in sentences (e.g., *I have seen that movie with my brother last week.). Results showed that participants in the distributed group were able to detect the grammatical errors better than those in the massed practice in the delayed post-test. Consistent with findings in cognitive psychology, research on discrete-point vocabulary and grammar learning indicate the results in favor for distributed practice over massed practice. The targeted skills and knowledge in these studies, however, underrepresent the complexity involved in L2 learning.

As Bird admits, the findings in his study cannot be generalized straightforwardly to L2 grammar learning in the broader sense. The practice and tests only entailed error correction exercises, i.e., the ability to make corrections to grammatical errors in the verb forms on a paper-and-pencil test. This is far from actual language learning and use. The current study aims to extend the notion of practice from the error correction exercises to a more meaning-focused oral practice. More meaning-focused practice is more aligned with the current practice of L2 teaching and research (DeKeyser, 2007). Furthermore, oral production requires a number of different processes (e.g., conceptualization and sound articulation) from error correction activities, introducing more complexity in grammar learning.

Complexity of skills to be acquired has been found to influence the effects of distributed practice. Donovan and Radosevich's (1999) review paper reported that distributed practice effects were attenuated in more complex tasks (defined by the degree to which the task requires a number of behaviors, choices, and uncertainty involved in the performance of the task). In their taxonomy, based on 95 university students' ratings, L2 vocabulary learning[2] was considered to have 'low' mental requirements and be of 'middle' complexity. The interesting question raised here is how L2 grammar practice of higher complexity influences the effects of distributed practice. The current study addresses this issue by using oral production practice for proceduralizing Japanese morphological structures.

II The present study

Following Bird's study, the present study aims to investigate the issue of how to distribute practice on discrete grammar points. It addresses whether distributed practice is more effective than massed practice for the proceduralization of an L2 morphological feature in oral production grammar practice. We extend the notion of practice from the mechanical written practice employed in Bird (2010) to a more meaning-focused oral practice. While Bird (2010) examined the acquisition of semantics of tense-aspect forms in English, the current study focuses on morphological markings on verbs that express the present progressive—six (phonologically determined) allomorphs of a verb in Japanese. This construction is suited for practicing oral production, since the meaning can be easily expressed in pictures and videos.

Based on Bird's findings that distributed practice was more effective than massed practice for L2 grammar learning, one can expect to find a similar pattern in the current study. Alternatively, since this study employs L2 training tasks that involve more diverse learning processes, the distributed practice effects might be attenuated due to the complexity of these processes, supporting Donovan and Radosevich's findings (1999).

1 Design

The study employed a pretest-posttest design with two experimental conditions differing in spacing of practice. Participants were randomly assigned to a 1-day ISI or a 7-day ISI group. As shown in Figure 1, each participant engaged in four individual sessions in a quiet laboratory: the pretests and training session 1 (Time 1), training session 2 followed by posttests (Time 2), posttests (Time 3), and posttests (Time 4). The tests were administered to measure knowledge of the target grammatical structure at Time 1 through Time 4. At Time 1, the tests were conducted before the training session to measure the pre-existing knowledge of the vocabulary and the target structure that were going to be practiced in the training. At

Time 2, the tests were administered immediately after training session 2 in order to assess how much participants had learned through practice. These tests were administered again at Time 3 and Time 4, which occurred 7 days and 28 days after Time 2, to assess retention of knowledge.

<table>
<tr><td></td><td>Time 1</td><td></td><td>Time 2</td><td></td><td>Time 3</td><td></td><td>Time 4</td></tr>
<tr><td>1-day ISI</td><td rowspan="2">Test
+
Training Session 1</td><td>→
1 day</td><td rowspan="2">Training Session 2
+
Test</td><td rowspan="2">→
7 days</td><td rowspan="2">Test</td><td rowspan="2">→
21 days</td><td rowspan="2">Test</td></tr>
<tr><td>7-day ISI</td><td>→
7 days</td></tr>
</table>

Figure 1. Research design

The ISIs were determined based on Rohrer and Pashler’s optimal ratio of ISI and test delay (RI: Retention Interval). As shown in Table 1, the ISI and RI were determined such that only one of the groups fell within Rohrer and Pashler’s optimal ratio of ISI and RI (10% to 30%) at the 7-day RI and the 28-day RI. Participants in the 1-day condition are expected to outperform those in the 7-day ISI condition at Time 3 (7-day RI) because the ratio of ISI to the testing is within the optimal range (14% vs. 100%). However, participants in the 7-day ISI condition are expected to perform better than those in the 1-day ISI condition at Time 4 in light of the optimal range of ratio (25% vs. 3%). It is important to note that the current study only included two practice sessions, whereas there were five training sessions in Bird (2010). This difference in research design will be discussed in detail later.

Table 1. Ratios of Intersession Intervals (ISI) to Retention Interval (RI)

	7-day RI (Time 3)	28-day RI (Time 4)
1-day ISI	14%	3%
7-day ISI	100%	25%

2 Target structure

The present study targeted a morphological structure in Japanese, the *te*-form of the verb, denoting a realized state or activity. It is used in a number of sentence constructions, and the study particularly focused on the usage of the present progressive, as in *-te imasu* (e.g., *Taro wa ki o nobotte imasu*; Taro-subject tree-object is climbing). There are consonant-ending-stem verbs and vowel-ending-stem verbs in the Japanese regular verb paradigm. There is no change in the stem of the vowel verbs for the *te*-form, whereas consonant verbs involve an allomorphic stem change. The focus of the study is on six groups of consonant verbs as shown in Table 2. Three verbs were taken from each of six categories for the training session (see Appendix 1 in the online supplementary material for all 18 verbs). All the verbs were action verbs and were unknown to participants as shown by the pretest scores.

Table 2. Conjugation of the *Te*-Form

Stem	Transformation Rule	Uninflected form	te-form
- r	Q	*nobor-u* (to climb)	*nobot-te*
- vowel	Q	*hiro-u* (to pick up)	*hirot-te*
- m	n	*tatam-u* (to fold)	*tatan-de*
- b	n	*musub-u* (to tie)	*musun-de*
- k	i	*migak-u(*to polish)	*migai-te*
- g	i	*sosog-u* (to pour)	*sosoi-de*

The uninflected form[3] of these verbs is converted to the *te*-form with either the /te/ or the /de/ allomorph. When the stem ending of a consonant verb is /r/ or /w/,[4] it turns into /Q/[5] in the *te*-form (e.g., *nobotte* for *noboru* or *hirotte* for *hirou*); when /the ending is m/ or /b/, it turns into /n/ (e.g., *tatande* for *tatamu* or *musunde* for *musubu*); and when it is /k/ or /g/, into /i/ (e.g., *migaite* for *migaku* or *sosoide* for *sosogu*) (Vance, 1987).

3 Participants

Forty beginner-level learners of Japanese as a second language participated in the study (25 female and 15 male). Their mean age was 21 years old (*SD* = 2.89). They all were compensated 45 US dollars for their participation. We distributed flyers across campus at a mid-Atlantic university and recruited 29 participants who were currently enrolled in third-semester Japanese courses during the study. In order to secure enough participants for statistical analyses, we further recruited 11 participants who had taken Japanese courses for two semesters before but were not taking any Japanese courses at the time of study. These two groups of participants will be analyzed separately as well as the group as a whole (See Coding and analysis section). The target structure, the present progressive, is introduced in the second semester of Japanese, and all the participants had learned about the *–te* form before the study, but most of them did not master it completely.

The participants were randomly assigned to one of the two treatments: 18 participants in the 1-day ISI condition and 22 participants in the 7-day ISI condition.[6] Of the 29 participants who were enrolled in third semester Japanese courses during the study, 13 were assigned to the 1-day ISI group and 16 were assigned to the 7-day ISI group. Of the 11 participants who were not taking any Japanese courses, five were assigned to the 1-day ISI group and six were assigned to the 7-day ISI group.

The first language of the participants was English except for two individuals (their first languages were Nepali and Romanian), but they were included in the study because these two participants were highly proficient, using English for their undergraduate study.

4 Outcome measures

In the present study, knowledge of the *–te* form was assessed in two ways: accuracy and cognitive fluency. Accuracy indicates to what extent participants can use the *te*-form appropriately in production. Cognitive fluency is part of the taxonomy of fluency proposed by Segalowitz (2010) and was defined as an aspect of procedural/automatized knowledge. It refers to the efficiency of integration and execution of the operations involving speech planning, lexical search, grammatical encoding, and articulation (see also De Jong, Steinel, Florijn, Schoonen, & Hulstijn, 2013). Cognitive fluency is operationalized in the current study as the speed with which linguistic knowledge can be used. In order to measure proceduralization of linguistic knowledge, two types of oral tasks were employed as the pretest and posttests. The tests were computerized and administered with the DMDX software (Forster & Forster, 2003), and the responses were audio-recorded.

Rule application test. The purpose of the rule application test was to assess the degree of proceduralization of the *te*-form rules. Eighteen nonce verbs were created based on the practiced verbs, keeping the same initial phoneme and number of moras, and different nonce verbs were used in each test to avoid practice effects (see Appendix 2 in the online supplementary material). Participants were required to convert the sentence with an uninflected verb into the one with present progressive by using a pseudo-Japanese (nonce) verb; they were told, however, that they would see and hear a “trendy” Japanese verb. After a fixation cross appeared in the center of the screen, an uninflected form of a nonce verb (e.g., *madasu*) was presented both in the written and oral modality. Immediately after the end of the word, participants were asked to convert the word into the present progressive form as quickly as possible (e.g., *madashite imasu*). They were given 10 seconds to complete their response. Participants performed a practice session with five unknown pseudo verbs in order to become (re)familiarized with the format of the test every time they took it. It took approximately three minutes to complete the test.

Picture sentence completion test. As the rule application test only targeted the rules for the *te*-form, the picture sentence completion test assessed to what extent the participants could use the correct te-form of the verbs that they practiced. In the picture sentence completion test, participants were presented with a picture in which someone was performing an action, immediately followed by an auditory stimulus, which was the subject of a sentence to be completed (i.e., *otokonohito* (man), *onnanohito* (woman), *otokonoko* (boy), or *onnanoko* (girl)). Their task was to complete the sentence by describing what the person was doing (e.g., *ki o nobotte imasu*; [he/she] is climbing the tree). They were given a maximum of 15 seconds to complete their responses. Four practice items with two basic verbs (i.e., eat a hamburger and watch TV) preceded the actual eighteen items. At the pretest, participants were shown pictures to check what each picture meant before they took the picture sentence completion test. It took approximately five minutes to complete the test.

5 Procedure

Each participant was engaged in all four sessions in a quiet laboratory. As shown in Figure 1, the two tests of *te*-form knowledge were administered before the training at Time 1 and after the training session at Time 2. The rule application test was always administered before the picture sentence completion task because the first test assesses narrower knowledge (rules) than the second (rule + vocabulary). In this order, a smaller practice effect, if any, was expected to occur from the narrow to the broad test, rather than vice versa. No feedback was given throughout the testing phase.

Each training session consisted of four tasks that helped learners practice the use of the present progressive in an explicit step-by-step manner. The training set took about 45-50 minutes. In the vocabulary learning task, 18 verbs (with object nouns) were learned on the computer screen. First, a picture was presented on the screen for five seconds. The participants were required to say the uninflected form of a verb phrase within this time period (e.g., *batto o huru*; swing the bat). Immediately after these five seconds, they were presented with the Japanese phrase both aurally and visually, in blue letters, along with the written equivalent of the English translation in black. The vocabulary remained on the screen for five seconds, and the next picture appeared automatically unless they proceeded by themselves. The vocabulary learning task was carried out individually, and an experimenter sat behind each participant and coded the answers. They repeated the set seven times, so that they could remember most of the words and use them in further practice.

Explicit grammatical explanation was provided after the vocabulary learning (see Appendix 3 in the online supplementary material). It was presented on a sheet of paper, and participants were told to read the explanations and to read aloud the verb in the conjugation chart. This sheet was available for reference during the entire practice session. In order to make sure that participants knew the correct conjugation forms, they were given a worksheet in which they transformed the uninflected form of the 18 verbs (accompanied by the pictures from the vocabulary learning task) into the present progressive *-te* form.

After the explicit explanation, participants completed the comprehension practice of sentences in the present progressive one on one with the experimenter. For this practice, cards that had the same pictures as the ones used during the vocabulary training session were laid out on the table. The experimenter read aloud the sentence that described the action in one of the pictures, and the participant's task was to pick up the corresponding card as soon as possible. This task was repeated twice.

The production task, also with picture matching, came next. The roles were reversed from those in the comprehension practice: Participants were asked to describe the picture to the experimenter, so that he could pick up the picture that participants described. When participants could not describe the picture, the experimenter described the card for them. This task was also repeated twice. Feedback in the form of recasting was given if participants produced an incorrect form of the verb. No time pressure was imposed in the picture-matching task to allow for careful rule application.

As a final task in the training set, participants performed a narrative task, describing what a person in a video was doing. Each action was performed for ten seconds, and the participants were told to describe the action using the *-te* form while the video was played. After each video clip, the correct sentence was presented both aurally and visually on the screen for four seconds, and the participants automatically moved on to the next movie clip. As in the case of the two previous tasks, the same video narrative task was performed twice to ensure that participants received enough practice on the sentences. All tasks were conducted with the procedure delineated earlier, both in training sessions 1 and 2.

6 Coding and analysis

Five trained independent raters, whose first language is Japanese, conducted analyses of the two outcome tests. Raters were trained to code accuracy of response and measure response durations using the sound analysis software Praat. The raters were trained until their coding matched those of the present researcher using 15% of data.

Rule application test.

First, accuracy of the utterances was coded by listening to each utterance. Minor pronunciation errors were ignored in the scoring, and the accuracy of utterances with repair was determined based on the last utterance. Reliability of the accuracy scores in the rule application test across time was calculated with Cronbach's alpha, and the indices were .941 at Time 1, .890 at Time 2, .892 at Time 3, and .916 at Time 4.

The participants' utterances were also analyzed for speed or Response Time (RT), measured from the onset of the prompt word to the end of the utterance.[7] RT of responses that contained incorrect utterances was excluded from the analysis. In addition, RT was not calculated for responses with repairs, rephrasing, and/or false starts, because it was impossible to determine whether they were due to lack of linguistic knowledge or other random sources (e.g., slip of the tongue).

In order to compute average RTs for each participant reliably, we set two further criteria for data exclusion. First, a certain number of correct utterances were needed to compute the average RT reliably for each participant. In the previous study that measured

RTs in a similar way, the cut-off point for accuracy was set at 65% (De Jong, 2005). We set the lower cut-off point at 33% (accuracy score of 6 or higher out of 18) and excluded participants whose accuracy scores were lower than the cut-off. This cut-off point was decided after the inspection of data in order to keep a larger number of participants while still having stable RTs for each participant. In total, the percentages of valid temporal measures retained were 29.7%, 83.2 %, 81%, and 80% for Time 1 through Time 4, respectively. Given the small percentage of valid responses for Time 1, temporal measure of the rule application test at Time 1 were not included in the subsequent analyses.

Second, in order to exclude responses resulting from different processes than normal responses, we defined outliers as RTs below the minimum of 500 ms and RTs higher than 3SD above the grand mean for each participant. These cutoff values were determined after inspection of the data (see De Jong et al., 2013 for a similar approach). The number of outliers identified ranged from 3.7% to 8.7% of the data sets across the tests.

Picture sentence completion test.

Responses in the picture sentence completion test were analyzed in terms of accuracy and speed, as in the rule application test. Accuracy was scored based on the accuracy of the *te*-form, and minor pronunciation mistakes in vocabulary were ignored (e.g., *sukutte imasu* for *tsukutte imasu*). Reliability of the accuracy scores was calculated with Cronbach's alpha, and the indices were .801 at Time 2, .862 at Time 3, and .875 at Time 4.

Similarly to the rule application test, the speed was calculated from the onset of a picture to the end of the utterance. The percentage of outliers (falling outside the range of minimum 500ms, and maximum 3SD above the grand mean, as for the rule application test) ranged between 2.3% and 6.2% of the data sets across the tests. After excluding participants whose accuracy scores were 5 or lower, the percentages of valid RT data retained were 80 %, 62.8%, and 57.4% for Time 2 through Time 4. Note that as accuracy was 0 for Time 1, no valid RT was measured at Time 1.

Main analysis.

For the main analysis, a mixed ANOVA was conducted with time (Times 1, 2, 3 and 4) as within-subject factor and group (1-day ISI and 7-day ISI) as between-subjects factor. We had four dependent variables (accuracy and speed in both the rule application test and the picture sentence completion test), and we ran four separate mixed ANOVAs.[8] Since speed in the two tests was not calculated at Time 1, ANOVAs for RT did not include Time 1. If a significant interaction between group and time was detected in the mixed ANOVA, t-tests were also conducted for the gain score from Time 1 to Time 3 and for that from Time 1 to Time 4 (with a Bonferroni-corrected alpha level of 0.025).[9] Time 3 and Time 4 only were used to test a priori hypothesis regarding the ratio of ISI and RI (see Table 1). The effect sizes were computed for ANOVAs (a partial eta squared), using the following criteria: small ($\eta^2 = 0.01$), medium ($\eta^2 = 0.06$), and large ($\eta^2 = 0.14$); and for *t*-tests (Cohen's *d*), using as criteria: small ($d = 0.2$), medium ($d = 0.5$), and large ($d = 0.8$) (Cohen, 1988). Confidence Intervals (CIs) for the effect sizes were also computed.

As explained previously, the study sample consisted of two different groups: Japanese L2 learners who were currently taking courses ($n = 29$) and those who were not anymore, although both groups had learned the -*te* form in the second-semester Japanese course (n = 11). The majority of participants were enrolled in third-semester Japanese courses at the time of study, and we also conducted the same analyses separately for this group only (13 participants in the 1-day ISI group and 16 participants in the 7-day ISI group). This subset analysis, albeit with a smaller sample size, improves the internal validity of the study by controlling confounding variables that could have (unexpectedly) existed between the two experimental groups. The other subset of participants were not analyzed due to the small number of participants. Where the subset analysis and the whole-group analysis produced different results, we presented the results of the subset analysis in greater detail to examine the discrepancies.

III Results

Descriptive statistics for all the measures are available in the online supplementary material (see Appendices 4, 5, 6 and 7). We first checked that the performance at Time 1 was equivalent for both groups. At Time 1, the mean scores on the rule application test were 5.06 (*SD* = 5.42) in the 1-day ISI group and 5.91 (*SD*= 5.55) in the 7-day IS group. According to Levene's test, the assumption of equality of variances was met ($p = .34$). No significant difference was detected between the two conditions, $t\ (38) = -0.49$, $p = .63$. For the subset analysis, the mean scores were 6.54 (*SD* = 5.53) in the 1-day ISI group and 7.81 (*SD* = 5.29) in the 7-day ISI group. The assumption of equality of variances also met ($p = .84$) here, and no significant difference was found between conditions here either, $t\ (27) = -0.63$, $p = .53$. For the picture sentence completion test, none of the participants knew any of the words, that is, no difference existed between groups at Time 1.

It could be also argued that the participants in the two groups were different in terms of the amount of time they devoted to learning Japanese outside of the classroom. When we asked them, however, how many hours they studied Japanese outside of the classroom between Time 1, Time 2, Time 3, and Time 4, the average numbers of study hours reported were 1.33 (*SD* = 1.85) and 0.89 (*SD* = 0.83) at Time 2[10], 4.83 (*SD* = 7.02) and 5.96 (*SD* = 5.79) at Time 3, and 11.84 (*SD* = 19.20) and 13.91 (*SD* = 12.18) at Time 4, for the 1-day ISI and the 7-day ISI group respectively. None of the differences were statistically significant by t-tests ($p < .05$), so the findings are not affected by different amounts of practice or effort. After confirming no existing difference between the groups, results are presented first for accuracy measures on both tests, followed by speed measures. Results for the whole group and subset are also presented side by side.

1 Accuracy measures

Figure 2 plots the mean accuracy scores with error bars (representing standard error of the mean) for the two outcome tests. The general trend for accuracy in both tests is that the massed practice group outperformed the distributed practice group across times after the treatment, and the gap was becoming smaller at delayed tests. The same patterns were also found for subset analysis (Figure 3).

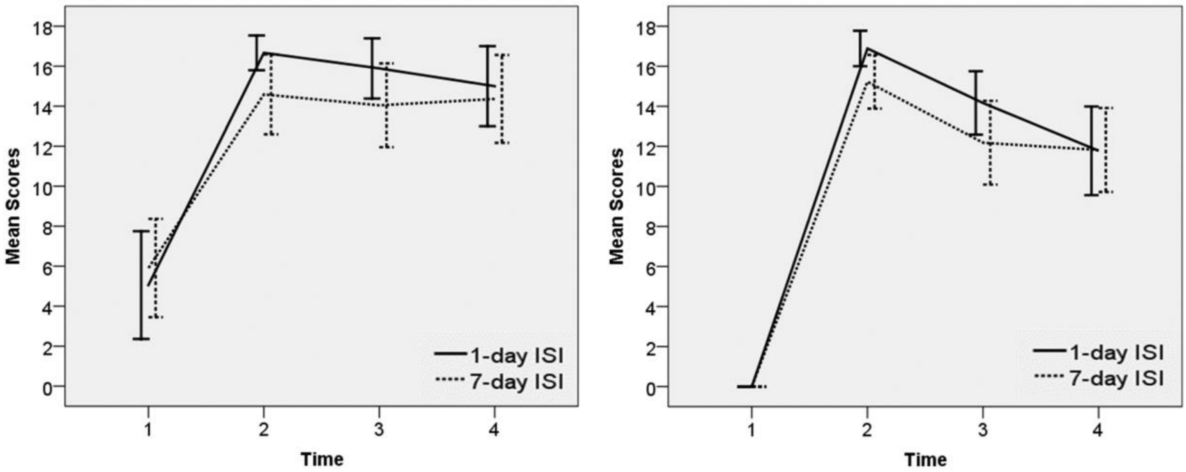

Figure 2. Accuracy scores across time for the whole group: the generalization test (left) and the sentence completion test (right).

Note. The error bars indicate 95% CIs.

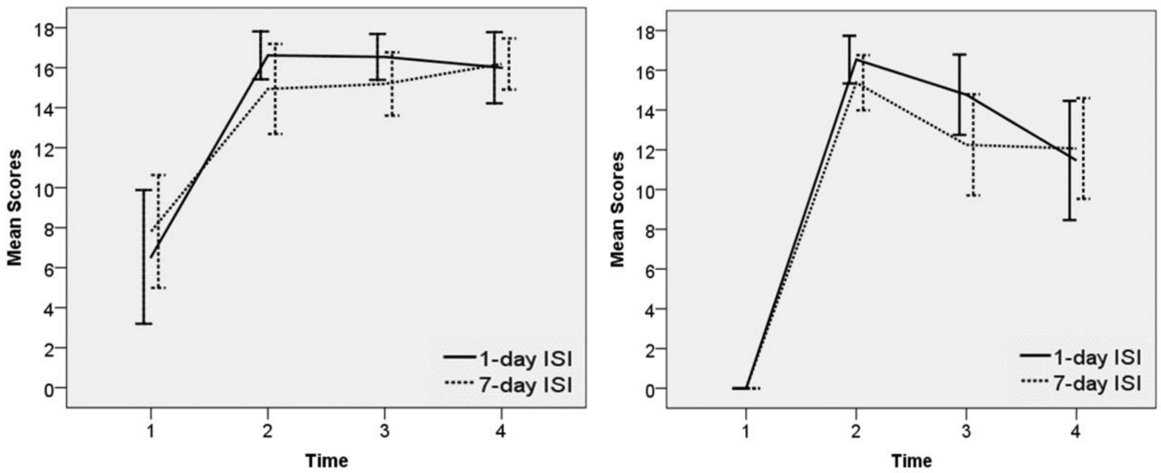

Figure 3. Accuracy Scores Across Time for the Subset: the Generalization Test (left) and the Sentence Completion Test (right).

Note. The error bars indicate 95% CIs.

A mixed ANOVA was first conducted on accuracy in the rule application test. Mauchly's Test of Sphericity indicated a violation of the sphericity assumption, $x^2(5) = 0.37$, $p < .001$; we therefore report the results using the Greenhouse-Geisser correction. There was a significant main effect for Time, F (1.97, 74.93) = 98.04, $p < .001$, $\eta^2 = 0.72$, 90% CI [0.63, 0.77].[11] There was no significant main effect for group, F (1, 38) = 0.66, $p = .42$, $\eta^2 = 0.02$, 90% CI [0.00, 0.13], nor was there a significant interaction between time and group, F (1.97, 74.93) = 1.90, $p = .16$, $\eta^2 = 0.05$, 90% CI [0.00, 0.13]. Although the score in the 1-day ISI is always higher after the treatment (Time 2 through Time 4), the effects of the two treatments on accuracy seemed to be comparable. The same analysis was conducted with the subset (participants who were enrolled in third-semester course), and the same results were obtained (see also Figure 3).

A mixed ANOVA was also conducted on accuracy in the picture sentence completion test. Mauchly's Test of Sphericity indicated violation of the sphericity assumption, $x^2(5) = 414$, $p < .001$, and the mixed ANOVA with the Greenhouse-Geisser correction revealed a significant main effect for time, F (1.89, 71.67) = 3144.27, $p < .001$, $\eta^2 = 0.99$, 90% CI [0.98, 0.99]. Similarly to the pattern of the accuracy scores in the rule application test, accuracy in the picture sentence completion test was also higher in the 1-day ISI group in the beginning, but the gap closes at Time 4. There was no significant main effect for group, F (1, 38) = 1.21, $p = .28$, $\eta^2 = 0.03$, 90% CI [0.00, 0.16], or interaction between time and group, F (1.89, 71.67) = 1.91, $p = .16$, $\eta^2 = 0.05$, 90% CI [0.00, 0.13]. The subset analysis showed the same pattern as the whole-group analysis (see also Figure 3).

2 Speed measures

The speed measures in the outcome tests showed somewhat different patterns from the accuracy measures. As shown in Figure 4, although the 7-day ISI group seems to respond faster than the 1-day ISI group in the rule application test, the differences appear to be marginal as the error bars (indicating the standard error) are overlapping across time. In contrast, the speed in the sentence completion test seems faster in the 1-day ISI group, particularly at Time 4 (i.e., the difference is about 600 milliseconds). When comparing the plots in Figure 4 (whole group) and in Figure 5 (subset group), the rule application test shows the opposite pattern in the subset analysis; the 1-day ISI group seems to respond faster than the 7-day ISI group across time, and the difference seems smaller at Time 4. The speed measures in the sentence completion test are almost identical for the subset analysis; the 1-day ISI group outperforms particularly at Time 4.

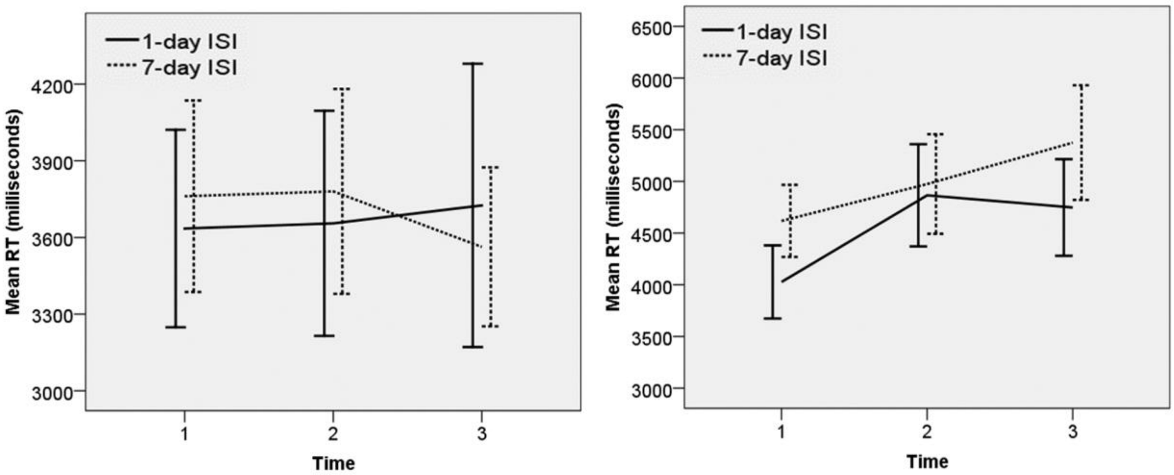

Figure 4. Speed measures across time for the whole group: the generalization test (left) and the sentence completion test (right)

Note. The error bars indicate 95% CIs.

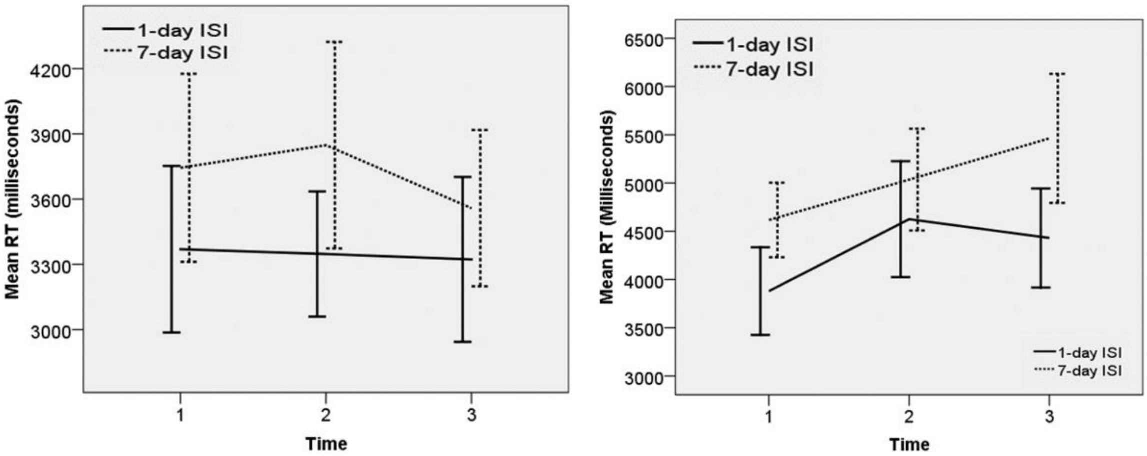

Figure 5. Speed measures across time for the subset: the generalization test (left) and the sentence completion test (right)

Note. The error bars indicate 95% CIs.

A mixed ANOVA was conducted on speed in the rule application test. Mauchly's Test of Sphericity indicated that the assumption of sphericity was met, $x^2(2) = 0.97$, $p = .64$. No main or interaction effects were significant: time, F (2, 68) = 0.50, $p = .61$, $\eta^2 = .02$, 90% CI [0.00, 0.07], group, F (1, 34) = 0.22, $p = .72$, $\eta^2 = 0.00$, 90% CI [0.00, 0.11], time*group, F (2, 68) = 2.14, $p = .13$, $\eta^2 = 0.06$, 90% CI [0.00, 0.15]. This suggests that the performance between the two groups was comparable. Although the subset group analysis showed a different pattern from the whole group analysis at the descriptive level (Figure 5), the results of the ANOVA were basically the same.

A mixed ANOVA was also conducted for the speed measure in the picture sentence completion test. Mauchly's Test of Sphericity indicated that the assumption of sphericity was met, x^2 (2) = 0.97, $p = .61$. The ANOVA revealed a significant interaction between time and group, F (2, 68) = 3.75, $p = .03$, $\eta^2 = .10$, 90% CI [0.01, 0.20], as well as a significant main effect for time, F (2, 68) = 36.09, $p < .001$, $\eta^2 = 0.52$, 90% CI [0.36, 0.61]. No significant main effect for group was found, F (1, 34) = 1.55, $p = .22$, $\eta^2 = 0.04$, 90% CI [0.00, 0.19]. The significant interaction indicates that improvement in speed was mediated by group, and t-tests were conducted for Time 3 and Time 4. According to Levene's test, the assumption of homogeneity of variances was met at both Time 3 ($p = .95$) and Time 4 ($p = .59$). There was no significant difference between the groups at Time 3, t (35) = -0.33, $p = .74$, 95% CIs of the difference [-556, 775], $d = 0.11$, 95% CI of d [-0.54, 0.76], but a marginally significant difference was found at Time 4 with a medium effect size, t (35) = -1.81, $p = .08$, 95% CIs of the difference [-76, 1333], $d = 0.59$, 95% CI of d [-0.07, 1.27].

The subset analysis showed a similar pattern; it just made the advantage of the 1-day ISI group more salient. The mixed ANOVA revealed a marginally significant main effect for a group with a large effect size, F (1, 24) = 3.89, $p = .06$, $\eta^2 = 0.14$, 90% CI [0.00, 0.34], as well as the significant main effect for time, F (2, 48) = 21.68, $p < .001$, $\eta^2 = 0.48$, 90% CI [0.28, 0.58], and the significant interaction, F (1, 48) = 3.90, $p = .03$, $\eta^2 = 0.14$, 90% CI [0.00, 0.21]. A set of t-tests showed that there was no significant difference at Time 3, $t(25) = -1.12$, $p = .28$, 95% CI of the difference [-346, 1166], $d = 0.43$, 95% CI of d [-0.34, 1.19], and the difference at Time 4 was significant (with a Bonferroni correction) with a large effect size, $t(25) = -2.63$, $p = .02$, , 95% CI of the difference [223, 1844], $d = 1.01$, 95% CI of d [0.20, 1.81].

IV Discussion

The current study investigated distributed practice works better than massed practice in L2 grammar learning. It addressed whether L2 learners still benefit more from distributed practice than massed practice even if we expand the notion of practice from the grammatical error correction tasks employed in Bird (2010) to a more meaning-focused oral production task. Contrary to what was the case in Bird's study, there were no statistically significant differences between the two groups in accuracy scores, for neither of the two tests. The participants in the massed group outperformed the distributed practice group at Time 2 and Time 3 at the descriptive level, but there was a trend suggesting that accuracy scores in the distributed practice group were approaching those in the massed practice group at Time 4 (i.e., the 28-day delayed test). One could argue that distributed practice might have yielded higher scores than massed practice on an even more delayed test (e.g., 45 days or 60 days RI), at least for accuracy. This is a possible scenario, but the ratio of ISI and RI at Time 4 in the study was almost identical to the most delayed test in Bird (2010): the ratios of the massed group were 3% and 5 % in the current study and Bird's, and those of distributed practice group were 23% and 25%, respectively. Distributed practice does not seem to have the same advantages in the present study as in Bird's.

On the contrary, the results of the speed measures even show an advantage of massed practice in the picture sentence completion test. In the whole group analysis, the massed practice group outperformed the other at Time 4 in terms of speed in the picture sentence completion test (with a medium effect size). The more controlled subset group analysis showed a significant difference with a large effect size for the comparison at Time 4, and the group difference overall was approaching significance, with a large effect size. These advantages for the massed practice group should be taken very cautiously because the CIs of effect size varied to a great extent. The 95% CI for Cohen's *d* in the whole group analysis ranged from no effect to a large effect, [-0.07, 1.27], and in the subset analysis, the effects varied from small to large, [0.20, 1.81]. Future research needs to be conducted with a bigger sample size in order to gain more confidence in the present findings.

In what follows, we first offer two related accounts that can explain the current findings in comparison with those in Bird's study: complexity and proceduralization of grammar learning. Next, the vocabulary learning processes in the training sessions were closely examined to seek potential explanations for the advantage in speed measures of the picture sentence completion task for the massed practice group. After that, we introduce two other factors (frequency of practice sessions/ISIs and ratios of ISI and RI) that should be carefully explored for future research along with the first two accounts. Instead of determining the only one factor that accounts for the findings, we think it is important to present multiple relevant factors to stimulate this area of research.

1 Higher complexity might moderate the distributed practice effects

One critical difference between the present study and Bird's might come down to the complexity of training tasks and outcome tests. As suggested in Donovan and Radosevich (1999), "overall complexity, mental, and physical requirements" (p. 798) influence the distributed practice effects. The current training tasks and outcome measures required more diverse cognitive processes than the paper-and-pencil error correction test in Bird's study. Here participants had to execute several speech processes (Levelt, 1989): conceptualize the meaning of present progressive, retrieve lexical information, apply a morphophonological rule to the verb, and articulate the sound. This higher complexity of the tasks and tests required in the present study probably attenuated the distributed practice effect (i.e., no differences between distributed and massed practice).[12]

Related to the complexity accounts, another important difference between Bird's study and the present one might lie in the difference in skill learning stage involved in the training (declarative knowledge versus procedural knowledge). The complexity accounts and this account are not mutually exclusive; each contributes to some aspect of the cognitive processes involved in distributed/massed learning. Learners in Bird (2010) engaged in error correction tasks, which presumably increased mainly declarative knowledge, with perhaps some incipient proceduralization for some learners. Since declarative knowledge is highly susceptible to memory decay, distributed learning may work better than massed practice in the initial stage of learning (i.e., declarative learning) (Kim, Ritter, & Koubek, 2013). In contrast, the present study focused on a later stage of skill acquisition, i.e., development of procedural knowledge and the first stages of automatization. The training and outcome tests in this study involved oral production tasks in a loosely timed setting, which required not only learning morphophonological rules but also the procedural knowledge to apply them quickly. This might have resulted in higher complexity of learning processes than error corrections tasks aimed mainly at the development of declarative knowledge.

Indeed, the most crucial difference was found for the speed measures, not accuracy, on the sentence completion test. A speed measure is a more sensitive index of procedural knowledge because it not only tests the accuracy of conjugation, but also how fast the morphological knowledge can be deployed. Furthermore, the picture sentence completion test required the most complex skills among the tests in the study because they required integration of lexicon and grammar. Our results, then, led us to a stronger interpretation of Donovan and Radosevich (1999), i.e. when learning and/or outcome tests become more complex, massed practice can not only be equally as effective, but even more effective than distributed practice. This interpretation is speculative but consistent with the findings in some macro-level foreign language classroom research showing that intensive courses work better than distributed courses. In those contexts, the learning involves far more complex tasks such as language comprehension and production in a variety of contexts, and the learning gains are usually assessed with tests requiring integration of different sources of knowledge.

In sum, second language learning aiming for one of the ultimate goals—acquisition of procedural or automatized knowledge—involves complex skills, and this complexity seems to have led to a different result in this study as compared to previous studies in SLA and psychology: massed practice can be more effective than distributed practice at the level of proceduralization or automatization.

2 Enhanced Lexical Retrieval in the Massed Practice Group

In addition to the accounts for the present findings above, the advantage for the massed practice group in speed measures on the picture sentence completion task was further examined. One of the reviewers pointed out that the learners in the massed practice group might have found it easier to retrieve vocabulary needed to describe the action in the picture sentence completion task and might therefore have been able to produce the sentences faster than those in the distributed practice.

In order to inspect the group differences for lexical knowledge, ad-hoc analyses were conducted on vocabulary learning phrases during the training sessions 1 and 2. Recall that each session had learners practice 7 sets of the same 18 uninflected verbs. Figure 6 presents the performance on vocabulary practice for the verbs.

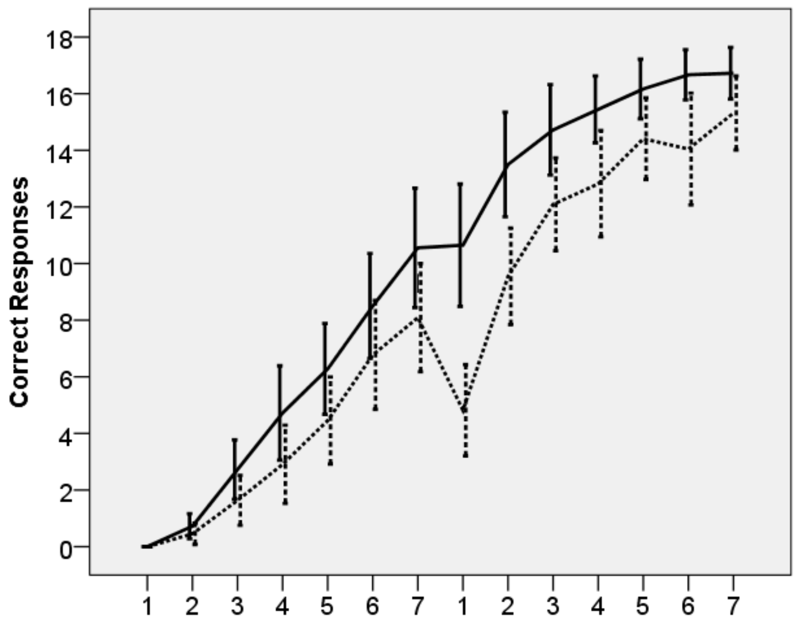

Figure 6. Correct responses during vocabulary practice in sessions 1 and 2

Note. The error bars indicate 95% CIs.

There were no significant differences between the two groups in any sets of training session 1. The critical difference between the two groups emerged from the very beginning of the training session 2. While the 1-day ISI group was able to maintain the accuracy of lexical retrieval from training session 1 to session 2, the 7-day ISI group showed a drop in accuracy in the first set of session 2. According to independent-samples t-tests, the massed practice group significantly produced more correct responses than the distributed practice group from the sets 1 to 6, *ps* $< .05$. No significant difference was found on the accuracy between the groups at the last set, $t(38) = 1.77$, $p = .09$. These results suggest that the learners in both groups underwent different learning processes in the second training session. The learners in the distributed practice group appeared to have suffered from recalling vocabulary in the earlier sets, whereas those in the massed practice group seemed to have more chances to practice the known vocabulary. The learners in the distributed practice had fewer opportunities to practice vocabulary for retrieving quickly, while the learners in the massed practice group were able to engage more effectively in practicing for retrieving lexical items more quickly. These different learning processes may account for the faster utterances on the picture sentence completion task. The stronger memory traces for the massed practice group during training session 2 may have led to the superior performance. If the lexical retrieval account was a primary explanatory factor for the present finding, then what the study tested can be considered to be lexical learning rather than morphological learning. It is, however, hard to distinguish lexical and morphological learning in practice; grammar practice is usually accompanied by vocabulary practice, and it is often unrealistic to ensure that all the vocabulary involved in a grammar-focused activity is already familiar to (and easily retrieved by) the students. In sum, the lexical retrieval factor seems a strong contributing factor for different cognitive processes in distributed and massed learning as well as the two accounts presented above. Given the novelty of the present findings, we have presented all the plausible interpretations; some may be favored over others based on theoretical interests.

3 Other factors: frequency of practice sessions and ratio of ISI and RI

In addition, the current study further shed light on two other factors concerning the distributed and massed practice. First, the difference between the current study and Bird's raises interesting issues of frequency of practice sessions and ISIs. The current study only employed two practice sessions (mainly due to practical reasons), whereas Bird (2010) incorporated 5 study sessions over the length of a 14-week semester. The third and fourth sessions in the present study involved posttests using the *te*-form, but no feedback was given at any point. This difference might have influenced the magnitude of the effect of distributed practice. Note that many studies in cognitive psychology included only two sessions (i.e, one ISI) to examine distributed practice effects, and the frequency of study sessions has not received focal attention (see recent review papers on distributed practice in Carpenter et al., 2012; Cepeda et al. 2006). The contrast between Bird's study and the present study may put a spotlight on this potentially important factor. In Bird's study, distributed practice effects were observed possibly because there were four repeated ISIs between the five training sessions. Distributed practice effects may be more likely to be observed for long-term retention of skills with more repeated practice and ISIs (Bird, 2010). A larger number of practice sessions and ISIs may consolidate the memory more strongly over time particularly when the practice is distributed.

Second, although the ratio of ISI and RI at Time 4 was almost comparable between the two studies, it was very different at Time 3. The ratio of ISI and RI in the distributed practice group (100%) was much further out of Rohrer and Pashler's optimal range (10%-30%) than the one in the massed practice group (3%). It may be the case that the disadvantageous ISI/RI ratio in the distributed practice group at the earlier point made it harder to observe the distributed practice effects. One of the reviewers, however, pointed out that when the ISIs are longer than the optimal ISI of 10-30%, the spacing tends to have lesser effects and that using the ISI/RI ratio of 100% at Time 3 was perhaps not a major issue for the distributed practice group.

There are a variety of ratios that researchers can set for an experiment, and it is worth paying attention to the ratio across a range of time span (e.g., based on prior research). In addition to the primary accounts of complexity and proceduralization we proposed, the two other factors (frequency of practice sessions/ISIs and ISI/RI ratios) we introduced here also need to be carefully explored in future research.

V Conclusions

The present study revealed that massed practice worked at least equally as well as distributed practice for acquisition of procedural/automatized knowledge or fluency development. Results furthermore suggested that massed practice is potentially more effective than distributed practice for proceduralization of production of sentences requiring retrieval of vocabulary and grammar. (A combination of) potential factors may account for the present findings; the present exploratory study hopes to spur further research in this area. These findings and interpretations should be further attested by a research design with a large number of subjects and different types of grammar practice (e.g., meaning-focused tasks with less focus on grammar target structures) as well as a larger number of practice sessions.

VI Pedagogical implications

With the limitations above in mind, we would like to offer tentative pedagogical implications from the current study and Bird's. They are tentative because very few empirical studies have been conducted to examine the optimal interval for grammar learning. As Bird's study suggested, distributed practice may be more advantageous than massed practice for an earlier stage of skill acquisition. Learning materials to foster declarative knowledge may work better if distributed systematically, e.g., according to the ISI/ RI ratio. On the other hand, what the current study found is that massed practice may be as beneficial as distributed practice or even more in a later stage of skill learning. Although the advantage of massed practice could not be captured with high confidence or precision in the current study (see CIs of Cohen's *d* on the picture sentence completion test at Time 4), it can be a viable option to condense training within a brief period of time when one aims to maximize the effectiveness of the training for proceduralization and automatization.

Note, however, that the current study's findings cannot be translated directly to program-level classroom research because the current study dealt with only two practice sessions and one interval. In addition, the current study was laboratory-based, and this could also limit the generalization of the current findings to a regular classroom context. The present findings may have more direct implications, however, when applied to computer-assisted language learning that is accompanied by regular classroom instruction or extracurricular practice. Further research is needed that compares the effectiveness of distributed practice and massed practice in more traditional classroom settings.

Acknowledgements

(Blinded for review)

Declaration of conflicting interest

The authors declare that there is no conflict of interest.

Notes

[1] These studies are among the most strictly controlled studies in the macro-level foreign language classroom studies. The total learning time between experimental groups was at least well controlled. However, studies in the field often suffer from some limitations including but not limited to different total learning time between groups and lack of delayed posttests (Rohrer, 2015).

[2] It was not mentioned explicitly that foreign language learning was vocabulary learning, but it is evident that the existing literature is limited to vocabulary in psychology laboratory experiments.

[3] Japanese verbs are always inflected, so the uninflected verbs are technically non-past form. We called them uninflected verbs to make it clear that they are the original verbs to be conjugated.

[4] A verb in this category (e.g., hirou) ends with vowel, but it is assumed that there is a covert /w/ (e.g., hiro(w)u).

[5] /Q/ indicates the geminate or double consonant in Japanese, such as /tte/ or /Qte/ as in nobotte or /noboQte/.

[6] Despite the random assignment, the participants' schedules made the 7-day ISI group larger than the 1-day ISI group at the end of the study.

[7] The latency (the RT before the utterance started) data were also coded for both tasks. Results essentially showed the same pattern of results for the speed measures (i.e., combined RT of latency and utterance duration), but the effects were smaller.

[8] A Multivariate Analysis of Variance (MANOVA) was not conducted because each test has different number of participants, leaving few participants for the test.

[9] The accuracy on the picture sentence completion test was zero at Time 1. The speed was not computed on both tests at Time 1; the gain scores were thus equivalent to the posttest scores for these three measures at Times 2, 3 and 4.

[10] Since there was a 7-day interval between Time 1 and Time 2 for the 7-day ISI group, the total number of study hours divided by seven is reported here.

[11] Instead of 95% CIs, 90% CIs were provided for the F test because it is one-sided (Steiger, 2004).

[12] As pointed out by one of the reviewers, Bird's tasks might be more complex than the present tasks in some dimensions. In Bird (2010), learners had to discriminate similar form-function mappings of English tense systems, which can be more complex than using the only one target morphological feature in the current study. Bird's task could have been more challenging in interpreting tense/aspect. This can be a factor that may mediate the distributed practice effect, which needs further investigations.

References

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review, 111*(4), 1036-1060.
Bird, S. (2010). Effects of distributed practice on the acquisition of second language English syntax. *Applied Psycholinguistics, 31*(4), 635-650.
Bloom, K. C., & Shuell, T. J. (1981). Effects of massed and distributed practice on the learning and retention of second-language vocabulary. *The Journal of Educational Research*, 245-248.
Carpenter, S. K., Cepeda, N. J., Rohrer, D., Kang, S. H. K., & Pashler, H. (2012). Using spacing to enhance diverse forms of learning: Review of recent research and implications for instruction. *Educational Psychology Review, 24*, 369–378.
Carpenter, S. K., Pashler, H., & Cepeda, N. J. (2009). Using tests to enhance 8th grade students' retention of US history facts. *Applied Cognitive Psychology, 23*(6), 760-771.
Cepeda, N. J., Coburn, N., Rohrer, D., Wixted, J. T., Mozer, M. C., & Pashler, H. (2009). Optimizing Distributed Practice. *Experimental Psychology (formerly Zeitschrift für Experimentelle Psychologie), 56*(4), 236-246.
Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological bulletin, 132*(3), 354-380.
Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
Collins, L., & White, J. (2011). An intensive look at intensity and language learning. *Tesol Quarterly, 45*(1), 106-133.
De Jong, N. (2005). Can second language grammar be learned through listening. *Studies in second language acquisition, 27*(2), 205-234.
De Jong, N., Steinel, M. P., Florijn, A., Schoonen, R., & Hulstijn, J. (2013). Linguistic skills and speaking fluency in a second language. *Applied Psycholinguistics, 34*(5), 893-916.
DeKeyser, R. M. (2007). *Practice in a second language: Perspectives from applied linguistics and cognitive psychology*. New York, NY: Cambridge Universtiy Press.
DeKeyser, R. M. (2015). Skill acquisition theory. In B. VanPatten & J. Williams (Eds.), *Theories in Second Language Acquisition: An Introduction* (2nd ed., pp. 94-112). New York, NY: Routledge.
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*(5), 795-805.
Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers, 35*(1), 116-124.
Karpicke, J. D., & Bauernschmidt, A. (2011). Spaced retrieval: absolute spacing enhances learning regardless of relative spacing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(5), 1250-1257.
Kim, J. W., Ritter, F. E., & Koubek, R. J. (2013). An integrated theory for improved skill acquisition and retention in the three stages of learning. *Theoretical Issues in Ergonomics Science, 14*(1), 22-37.
Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: The MIT Press.
Long, M. H., & Doughty, C. J. (2009). *The handbook of language teaching*. Oxford, UK: Blackwell.

Nakata, T. (2015). EFFECTS OF EXPANDING AND EQUAL SPACING ON SECOND LANGUAGE VOCABULARY LEARNING. *Studies in second language acquisition, FirstView*, 1-35. doi: doi:10.1017/S0272263114000825

Pavlik, P. I., & Anderson, J. R. (2005). Practice and Forgetting Effects on Vocabulary Memory: An Activation‐Based Model of the Spacing Effect. *Cognitive science, 29*(4), 559-586.

Rohrer, D. (2015). Student Instruction Should Be Distributed Over Long Time Periods. *Educational Psychology Review*, 1-9.

Rohrer, D., & Pashler, H. (2007). Increasing retention without increasing study time. *Current directions in psychological science, 16*(4), 183-186.

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

Seabrook, R., Brown, G. D. A., & Solity, J. E. (2005). Distributed and massed practice: From laboratory to classroom. *Applied Cognitive Psychology, 19*(1), 107-122.

Segalowitz, N. S. (2010). *Cognitive bases of second language fluency*. NY: Taylor & Francis.

Serrano, R. (2011). The time factor in EFL classroom practice. *Language Learning, 61*(1), 117-145.

Serrano, R., & Muñoz, C. (2007). Same hours, different time distribution: Any difference in EFL? *System, 35*(3), 305-321.

Sobel, H. S., Cepeda, N. J., & Kapler, I. V. (2011). Spacing effects in real world classroom vocabulary learning. *Applied Cognitive Psychology, 25*(5), 763-767.

Steiger, J. H. (2004). Beyond the F test: Effect size confidence intervals and tests of close fit in the analysis of variance and contrast analysis. *Psychological methods, 9*(2), 164-182.

Toppino, T. C., & Gerbier, E. (2014). About practice: repetition, spacing, and abstraction. *The Psychology of Learning and Motivation, 60*, 113-189.

Vance, T. J. (1987). *An introduction to Japanese phonology*: State University of New York Press Albany.