Studies in Second Language Acquisition (2024), 1–25
doi:10.1017/S0272263124000251

RESEARCH ARTICLE

The effects of distributed practice on second language fluency development

Joe Kakitani[1,2] and Judit Kormos[1]

[1]Lancaster University and [2]Dokkyo Medical University
Corresponding author: Joe Kakitani; Email: i.kakitani@lancaster.ac.uk

(Received 23 April 2023; Revised 23 December 2023; Accepted 20 February 2024)

Abstract

This study examined the effects of distributed practice on second language (L2) speech fluency development. A total of 116 Japanese L2 learners of English were randomly divided into experimental or control conditions. Learners assigned to the experimental groups engaged in four fluency training sessions either in a short-spaced (1-day interval) or long-spaced (7-day interval) condition. Although different learning trajectories were observed during the training phase, the posttests conducted 7 and 28 days after the training showed similar fluency gains for the two groups, indicating that short- and long-spaced conditions were equally effective for developing L2 fluency. The current study extends the line of research in distributed practice and task repetition for L2 fluency development.

Introduction

In the field of education, a topic of great interest to both teachers and students is how they can make the best use of study time to maximize learning. A learning strategy that has been deemed effective in cognitive and educational psychology is distributed practice (Dunlosky et al., 2013). Research on practice distribution has a long history that dates back to the 19th century (Ebbinghaus, 1885), and its effects have been extensively examined (see Wiseheart et al., 2019). The general consensus of previous research is that knowledge is better retained when practice takes place in a distributed rather than massed fashion. In the past decade, there has been a surge of interest in distributed practice in second language (L2) research (S. K. Kim & Webb, 2022; Serrano, 2022). Most existing L2 studies, however, have focused on constructs of language knowledge such as vocabulary and grammar. Given that language learning also entails acquiring fluency in linguistic skills (e.g., reading, listening, writing, speaking), how distributed practice can be used to promote L2 skill acquisition is an important question from both theoretical and pedagogical perspectives (Y. Suzuki, 2023).

The current study extends the line of investigations into the effects of task repetition schedule on L2 fluency—a dimension of L2 performance that hinges highly on

© The Author(s), 2024. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/S0272263124000251 Published online by Cambridge University Press

procedural knowledge (Kormos, 2006; S. Suzuki & Kormos, 2023). Manipulating the timing of task repetitions has been shown to affect the fluency of the repeated performance (Bui et al., 2019), and the effects of practice schedule have been found to transfer to performance on a novel task (Y. Suzuki & Hanzawa, 2022). What is still relatively unclear in this line of research is the long-term implications of distributed practice on L2 fluency development. Research in cognitive psychology suggests that an ideal distribution of practice depends on the ratio of the interval between practice sessions (intersession interval [ISI]) to the interval between the final practice session and the time of testing (retention interval [RI]). However, no research to date has examined the effects of distributed practice on L2 fluency development by manipulating the ISI–RI ratio. To fill this research gap, the current study investigates the role of practice distribution in L2 fluency practice and development using a pretest–posttest–delayed posttest research design based on optimal ISI–RI ratios established based on research findings in cognitive psychology.

Literature review

Distributed practice research in cognitive psychology and second language acquisition

Massed practice refers to a condition in which practice takes place in an intensive manner, cramming the study time into a single session. Distributed practice or spaced practice, in contrast, refers to a condition in which study time is divided into multiple sessions by inserting some time or items in between. For example, practice can be distributed by either (a) studying item A, taking a break for some time (e.g., 3 minutes), and then practicing item A again, or (b) practicing item A, practicing other items (e.g., item B, item C), and then practicing item A again.

In cognitive and educational psychology, the effects of practice schedule have been extensively researched (Wiseheart et al., 2019). The general consensus of previous research is that long-term learning effects are greater when practice opportunities are spaced rather than massed. This phenomenon of the spaced condition producing superior learning effects is called *the spacing effect.* A similar line of research investigates the effects of two or more distributed practice conditions of varied spacing length as opposed to comparing distributed and massed conditions. This line of inquiry is considered more ecologically valid, as learning in the real-world context seldom takes place in a single session (Rohrer, 2015). The aim here is to examine *the lag effect*, which refers to the superior learning gains produced by longer spacing. *The distributed practice effect* is used as an umbrella term to refer to both spacing and lag effects (Toppino & Gerbier, 2014).

Although a wealth of studies supports the robustness of the distributed practice effect, the matter is not as simple as *having longer intervals is always better.* Previous research has shown that the ISI, or the interval between practice sessions, interacts with the RI, or the interval between the final practice session and the time of testing. Cepeda et al. (2008) examined the effects of different learning schedules on trivial fact learning in which multiple ISIs and RIs were systematically manipulated. The findings showed that for the RIs of 7, 35, 70, and 350 days, the ISIs of approximately 3, 8, 12, and 27 days led to optimal results for the recall test, corresponding to ISI–RI ratios of 43%, 23%, 17%, and 8%, respectively. For the recognition test, the optimal ISIs were approximately 1.6, 7, 10, and 25 days of the RIs, corresponding to ISI–RI ratios of 24%, 19%, 14%, and 7%, respectively. As a rule of thumb, the ISI–RI ratio of 10%–30% is referred to as an optimal range (cf. Rohrer & Pashler, 2007).

https://doi.org/10.1017/S0272263124000251 Published online by Cambridge University Press

A growing number of studies have examined the effects of distributed practice on L2 learning over the past decade. Much previous research has shown the superiority of spaced practice over massed practice (Koval, 2019; Miles, 2014; Yamagata et al., 2022), which is in line with the broader literature in cognitive psychology (Cepeda et al., 2006). In contrast to relatively consistent findings on spacing effects, studies investigating lag effects have painted an unclear picture, especially regarding the generalizability of practice schedule based on the optimal ISI–RI ratio to L2 learning; some studies have shown the superiority of longer over shorter spacing (Bird, 2010; Rogers, 2015), whereas others have failed to obtain similar findings (Kasprowicz et al., 2019; Y. Suzuki, 2017). One possible reason for the inconsistency across findings may be that the level of complexity involved in information processing varies across different L2 tasks (Donovan & Radosevich, 1999; Y. Suzuki et al., 2019). According to the study-phase retrieval theory (Thios & D'Agostino, 1976), retention of knowledge rests with successful and effortful retrieval of a previously learned item. In this view, longer spacing can be suboptimal when learning complexity is too high, as it can lead to retrieval failure. This is also consistent with the desirable difficulty framework (Bjork, 1994; Y. Suzuki et al., 2019), which posits that practice leads to optimal gains when the learning condition is difficult enough to induce maximal effort from the learners but not too difficult to make the retrieval unsuccessful. Greater difficulties can lead to initial decreases in performance during the training phase but might result in later increases in retention. For example, Bahrick and Hall (2005) conducted an experiment in which 41 undergraduate students in the United States learned and reviewed Swahili–English word pairs over four training sessions that took place in a massed, 1-day ISI, or 14-day ISI condition. The massed and 1-day ISI groups outperformed the 14-day ISI group during the training phase, but the posttest administered 14 days after the training showed that knowledge was best retained in the 14-day ISI condition. To better understand the relationship between complexity/difficulty and distributed practice effects, it is worth exploring the role of practice distribution in relation to L2 fluency, which involves highly complex mental processes that depend on procedural or automatized linguistic knowledge (Kormos, 2006; S. Suzuki & Kormos, 2023).

The role of task repetition in fluency development

Research in task-based language learning has yielded a substantial body of knowledge regarding the benefits of task repetition for L2 fluency development (Lambert et al., 2017; Sun & Révész, 2021). The positive impact of task repetition—defined as "repetition of a given configuration of purposes, and a set of content information" (Bygate, 2018, p. 2)—can be explained using Levelt's (1989) speech production model. In this model, speech production is viewed as a process consisting of three stages: conceptualization, formulation, and articulation. In conceptualization, speakers generate a preverbal message by activating the relevant concepts that reflect their communicative intention. Formulation converts the preverbal message into linguistic forms through various encoding processes (e.g., lexical, syntactic, phonological encoding). In articulation, the linguistic representations are transformed into audible sounds by moving the speech organs. Task repetition supports L2 speech production processes by reducing the speakers' cognitive load for conceptualization and formulation in the repeated performance, which leads to improvements in fluency (Lambert et al., 2017). Task repetition is beneficial also because it allows L2 speakers to reuse linguistic constructions (Y. Suzuki et al., 2022). That is, linguistic formulations required to complete the

https://doi.org/10.1017/S0272263124000251 Published online by Cambridge University Press

task are primed in the initial performance (e.g., lexical retrieval, syntactic construction), which are readily available for reuse in the repeated performance. Finally, task repetition can affect the development of automatic encoding procedures. According to the skill acquisition theory (DeKeyser, 2015), in explicit learning contexts, L2 knowledge such as vocabulary and grammatical rules is first encoded as declarative knowledge, which is the basis for acquiring procedural knowledge and eventually automatized knowledge. Task repetition is helpful in providing the practice that L2 learners need to proceduralize and automatize their linguistic knowledge that underlies L2 fluency (S. Suzuki & Kormos, 2023).

The degree to which these benefits are realized depends on the type of task repetition, which can be broadly categorized into *same-task* repetition and *procedural* repetition (or *task-type* repetition). In same-task repetition, learners engage in the exact same task with identical content and procedures. In procedural repetition, learners engage in different content but in the same procedural manner (e.g., changing the picture in picture-description tasks). Of the two types of task repetition, same-task repetition arguably has a greater impact on the repeated performance in terms of gains from the priming effects and reduction in the cognitive load for conceptualization/formulation, as it involves the repetition of the entire speech production process (i.e., conceptualization, formulation, and articulation). Research on fluency training (N. de Jong & Perfetti, 2011) suggests that same-task repetition is also beneficial for proceduralization of L2 speech processing when training involves 4/3/2 procedures (i.e., decreasing the time limit for task repetitions). However, research findings also indicate that procedural repetition might be as effective as same-task repetition for improving speech rate (Lambert et al., 2021), and even more beneficial for increasing syntactic complexity (Y. Kim & Tracy-Ventura, 2013). Furthermore, procedural repetition is potentially useful for generalizability (DeKeyser, 2018), as it yields variation in practice. In other words, by practicing under varying task contexts, learners might develop speaking skills that are transferrable to a new task (e.g., fluency transfer).

The role of practice distribution in fluency development

Although previous research has demonstrated the benefits of task repetition, little attention has been given to the issue of *when* to repeat the task (Rogers, 2023), especially within the domain of L2 fluency research. Bui et al. (2019) were the first to conduct a study specifically examining the effects of task repetition schedule on L2 oral performance. In their study, 71 L2 learners of English in Hong Kong performed a picture description task twice under five different spacing conditions (immediate, 1-day, 3-day, 1-week, and 2-week). The results showed that immediate repetition was most conducive to improving speech rate, whereas 1-week spacing led to the largest reduction of filled pauses and verbatim repetitions. These findings indicate that the amount of spacing does influence how fluently L2 speakers perform a task at the second enactment. However, because Bui et al. only examined the changes that occurred from Time 1 to Time 2 using an identical task, the extent to which the effects of task repetition schedules transfer to a new task was unknown.

To the best of our knowledge, only two studies to date have examined the effects of spacing intervals on L2 fluency development using a pretest–posttest research design. First, Kobayashi (2022) investigated the impact of spaced practice on performance transfer. In her study, 38 Japanese university learners participated in a short L2 speaking training intervention, performing the same picture narration task twice in

https://doi.org/10.1017/S0272263124000251 Published online by Cambridge University Press

either a massed or spaced (1-week) condition. The results revealed no statistically significant differences between the two groups during the training phase in terms of complexity, accuracy, lexical variety, or fluency; however, the posttest conducted 1 week later using a novel task showed a significant pretest–posttest increase in lexical variety for the spaced group. One limitation to this study is that the scope of fluency assessment was quite limited, as only a single index (pruned words per minute) was used.

Y. Suzuki and Hanzawa (2022) examined the impact of task repetition schedule on fluency transfer and *retention*. Using a pretest–posttest–delayed posttest research design, they conducted a quasi-experimental study with 79 Japanese university learners in intact classrooms. Four groups of students were assigned to either one of three experimental conditions (massed, short-spaced, or long-spaced) or a control condition. Those assigned to the experimental conditions engaged in fluency training, performing the same picture narration task six times with varied temporal spacing. The findings obtained by an immediate posttest showed that massed practice led to gains in breakdown fluency (fewer mid- and end-clause pauses) but adversely affected speed fluency (slower articulation rate) and repair fluency (more verbatim repetition). The delayed posttest conducted 1 week later did not show any statistically significant differences across the four groups. However, it is important to note that the spacing schedules adopted in this study were not based on the optimal ISI–RI ratio. For instance, the ISI–RI ratio for the delayed posttest in the long-spaced condition was 100% (7-day ISI/7-day RI), which falls well outside the optimal range (10%–30%; Cepeda et al., 2008).

Our review of the extant literature reveals that further research is required to elucidate the effects of distributed practice on L2 fluency development. More specifically, investigations into the generalizability of optimal ISI–RI ratios to fluency development deserve attention given that no empirical study has been conducted previously on this issue. Furthermore, as existing studies are quasi-experimental studies conducted in the classroom setting (Kobayashi, 2022; Y. Suzuki & Hanzawa, 2022), further investigations using randomized experiments are warranted to probe this topic with more rigorous experimental control, enabling causal inferences to be drawn.

Current study

The current study aims to address the deficiencies in the literature by investigating the effects of distributed practice on L2 fluency development using optimal ISI–RI ratios. Using a random assignment study based on a pretest–posttest–delayed posttest research design, the participants' fluency development was assessed longitudinally. Due to coronavirus disease 2019 restrictions, all the training and test sessions were held online using video conferencing software (Microsoft Teams or Zoom). Oral speech was recorded using the participants' own mobile device as well as the recording system on the video conferencing software for backup purposes. These procedures had been piloted before the outset of the study for feasibility. The following research questions guided the current study:

1) To what extent does distribution of practice (1-day ISI vs. 7-day ISI) influence L2 speakers' performance during the training phase?
   a) How does fluency change during each training session?
   b) How does fluency change over multiple training sessions?

https://doi.org/10.1017/S0272263124000251 Published online by Cambridge University Press

2) To what extent does distribution of practice contribute to L2 speakers' fluency development as measured by pretest–posttest–delayed posttest changes?

For research question 1, it was hypothesized that the 1-day ISI group would outperform the 7-day ISI group during each training session as well as over multiple training sessions. Considering that shorter-spaced practice is more intensive than longer-spaced practice (Serrano, 2011), we speculated that learners in the 1-day ISI group would improve their fluency more quickly compared with those in the 7-day ISI group during the training phase. This hypothesis is also based on the assumption of desirable difficulties that longer spacing leads to suboptimal performance during training (Bjork, 1994). Longer spacing is expected to induce higher cognitive demands in linguistic formulation; therefore, we anticipated that the 7-day ISI group would show diminished performance relative to the 1-day ISI group during the training phase.

Regarding research question 2, two possible scenarios were envisioned. First, it is possible that the 1-day ISI group would demonstrate greater fluency gains than would the 7-day ISI group in the first posttest, whereas the results would be the reverse for the delayed posttest (i.e., the 7-day ISI group outperforming the 1-day ISI group). This would be in line with previous research in cognitive psychology that has demonstrated distributed practice effects (A. S. N. Kim et al., 2019). The second possible scenario is that the 1-day ISI group would outperform the 7-day ISI group in both the first posttest and the delayed posttest. As fluency development is largely associated with proceduralization of speech processing (Kormos, 2006; S. Suzuki & Kormos, 2023), this second scenario would be in line with the assumption of the skill acquisition theory, namely, that procedural knowledge is better acquired through intensive practice and is more robust to decay (DeKeyser, 2015; J. W. Kim et al., 2013). This hypothesis is also based on the findings from a recent study that failed to observe the benefits of long-spaced practice on long-term fluency development (Y. Suzuki & Hanzawa, 2022).

Methods

Participants

One-hundred sixteen undergraduate L2 learners of English studying in a Japanese university (68% female, 32% male) were included in the analysis (mean age = 20.62; standard deviation [*SD*] = 1.21). An additional six students had initially signed up to participate in the study but were unable to complete all the sessions due to attrition; consequently, their data were excluded from all analyses. The participants were recruited based on the following criteria: They (i) are L1 Japanese speakers, (ii) are not currently enrolled in an English language course at the university, and (iii) do not use or speak English on a regular basis. The purpose of setting these criteria was to recruit a relatively homogenous sample of L2 learners to reduce the effects of individual differences, as well as to control for potential practice effects outside of the study. The students' average score ($M = 625.17$, $SD = 96.90$) on a standardized English test (Test of English for International Communication (TOEIC)) indicated that their English proficiency was approximately between A2 (elementary) and B1 (intermediate) in the Common European Framework of Reference. The number of participants was determined based on an a priori power analysis using G*Power (Faul et al., 2007). For a within-between 4 × 3 multivariate analysis of variance (four groups at three measurement points) with the effect size set to medium ($f = .25$) and power to .80, the power

https://doi.org/10.1017/S0272263124000251 Published online by Cambridge University Press

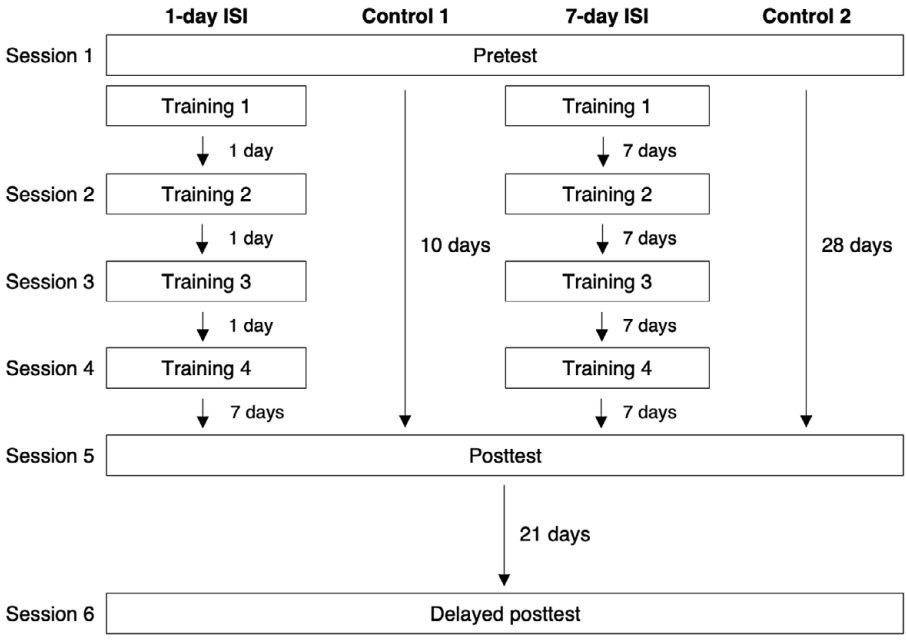

Figure 1. Research design.

analysis showed that a minimum of 113 participants would be necessary.[1] The total number of participants in the current study ($N = 116$) was thus considered sufficient.

Research design

Participants were randomly assigned to the 1-day ISI group ($n = 28$), control group 1 ($n = 29$), the 7-day ISI group ($n = 30$), or control group 2 ($n = 29$) (see Figure 1). One-way analysis of variance results indicated that there was no statistically significant difference across the four groups in terms of the TOEIC scores ($F[3, 112] = 0.01$, $p = .999$). The 1-day ISI and 7-day ISI groups engaged in four training sessions. The content and procedure of training sessions were identical for the two groups, with the only difference lying in the temporal spacing between the sessions. Namely, the participants assigned to the 1-day ISI condition practiced daily for 4 consecutive days, whereas those assigned to the 7-day ISI condition practiced once a week over 4 weeks. The participants assigned to the control groups, by contrast, did not participate in any training and only took the three tests (pretest, posttest, and delayed posttest), which followed the same schedule as each corresponding experimental

[1]Although we used generalized linear mixed-effects modeling (GLMM) for statistical analysis, we ran a priori power analysis for multivariate analysis of variance (MANOVA) using G*Power because there were no practical tools available for power analysis for GLMM at the time of study conceptualization (e.g., a summary-statistics-based power analysis; Murayama et al., 2022). MANOVA was chosen rather than ANOVA, because the current analysis involved multiple outcome measures and power analysis based on MANOVA produces a more conservative (larger) sample-size estimate.

https://doi.org/10.1017/S0272263124000251 Published online by Cambridge University Press

Table 1. Ratio of intersession interval (ISI) to retention interval (RI)

	7-day RI (posttest)	28-day RI (delayed posttest)
1–day ISI (short spacing)	14%	4%
7–day ISI (long spacing)	100%	25%

group (i.e., 1-day ISI group–control group 1; 7-day ISI group–control group 2). The results obtained from each control group provided a baseline for identifying any pretest–posttest changes that can be accounted for by the fluency training intervention.

The intervals between the training sessions (ISIs) and the intervals between the last training session and the posttests (RIs) were manipulated based on the optimal ISI–RI ratio suggested by previous research. Using the range of 10%–30% as a benchmark (Cepeda et al., 2008), we ensured that each experimental condition included an optimal condition at either the first posttest or the delayed posttest. Namely, the first posttest fell in the optimal range for the 1-day ISI group (14%), whereas the delayed posttest was in the optimal range for the 7-day ISI group (25%) (see Table 1).

Materials

Testing materials

Three picture prompts (*Bicycle*, *Race*, and *Soccer*) were used for the tests. They were six-frame cartoon stories adapted from Heaton (1966, 1975). The order of the prompts was counterbalanced to minimize task effects. All three prompts had a tight sequential structure (i.e., the picture frames are in a predetermined chronological order) and a storyline that encouraged speakers to express the feelings and motivation of the characters. Pilot testing confirmed that these picture prompts elicit similar oral performance from L2 speakers of English on measures of syntactic complexity, accuracy, and fluency (Kakitani, 2023).

Training materials

Four different picture prompts (*Hide-and-Seek*, *Picnic*, *Surprise*, and *Bus*) were used as training materials. They had a structure and storyline comparable to the testing materials. The prompts were presented to the participants in both experimental groups in the same order (*Hide-and-Seek* → *Picnic* → *Surprise* → *Bus*).

Procedure

Pretest and posttest sessions

Each participant individually joined the online sessions led by the first author or a trained research assistant using a video conferencing software. In the first session, participants signed an electronic consent form and answered a brief learner background questionnaire. Before administering the pretest and the posttests, instructions for the speaking task were given using a sample cartoon story, which was not part of the materials used for the tests or training. The participants were instructed to narrate the story so that even someone who has not seen the pictures could understand the story. They were given 3 minutes for planning and 4 minutes for narration in English. The use of a dictionary as

https://doi.org/10.1017/S0272263124000251 Published online by Cambridge University Press

well as notetaking were not allowed. The computer camera was on during each session to prevent potential nonadherence to instructions (e.g., taking notes, consulting a friend). The picture prompt was shown on the computer screen along with the title of the story written in English and Japanese. The participants were allowed to zoom in and out of the cartoon pictures during the planning time but not during the speaking time. A set of guiding questions was also provided in Japanese during the planning time to clarify the story and give the speakers additional ideas with regard to content (N. de Jong & Vercellotti, 2016). During the speaking time, the guiding questions were removed, and only the picture prompt was displayed on the screen.

Training sessions

The first training session took place on the same day as the pretest for the 1-day and 7-day ISI groups. No break was given between the pretest and the first training session. The training sessions followed a similar procedure as the test sessions described previously, with the time on task controlled for across the two experimental conditions. To remind the participants of the general procedure of the task, instructions on how to do the task were provided using a sample cartoon story at the beginning of each training session. One notable difference between the test and training sessions was the number of task repetitions. Namely, speakers narrated the story only once for the test, whereas they narrated the story three times during each training session (i.e., three task performances using the same picture prompt). Because previous research has shown the benefit of repeating the same task for enhancing L2 fluency (N. de Jong & Perfetti, 2011; Lambert et al., 2017; Y. Suzuki, 2021), the current study used same-task repetition in each training session to maximize the effects of oral practice. The 3-minute planning time was given before the first task performance but not between task repetitions; thus, the speakers completed three task repetitions without taking a break in between.

The participants engaged in four different narrative tasks over four training sessions (i.e., a different prompt was used for each training session). As discussed in the literature review section, procedural repetition (i.e., varying task contexts) is potentially more beneficial in enhancing generalization processes than same-task repetition (DeKeyser, 2018). Therefore, procedural repetition was deemed suitable for our study in which we aimed to examine the development of transferrable L2 speaking skills with the use of our posttest tasks. Furthermore, using different tasks across multiple training sessions is likely to be more ecologically valid and representative of classroom language learning contexts (e.g., a teacher varying tasks over several classroom sessions, aiming to practice the same skill or structure).[2]

Analysis

Data coding

Speech data were transcribed by trained research assistants based on analysis of speech units (Foster et al., 2000). Each transcribed text was subsequently double-

[2]The difference between the two experimental groups was in the spacing between procedural repetitions rather than same-task repetitions in the current study.

https://doi.org/10.1017/S0272263124000251 Published online by Cambridge University Press

Table 2. Fluency measures used to assess oral performance

Speed fluency
1. Articulation rate (mean number of syllables per minute, excluding pauses)
Breakdown fluency
2. Mid-clause pause duration (mean duration of silent pauses within clauses)
3. Mid-clause pause ratio (the number of within-clause silent pauses divided by the total number of syllables)
4. End-clause pause duration (mean duration of silent pauses between clauses)
5. End-clause pause ratio (the number of between-clause silent pauses divided by the total number of syllables)
6. Filled pause ratio (the number of filled pauses divided by the total number of syllables)
Repair fluency
7. Repetition ratio (the number of repetitions divided by the total number of syllables)
8. Self-repair ratio (the number of self-repairs divided by the total number of syllables)

checked by another research assistant and the first author to ensure accuracy. Following previous fluency research (S. Suzuki & Kormos, 2023), the transcribed texts were pruned by excluding filled pauses, repetitions, and self-repairs. One of the participants was unable to complete the delayed posttest on the scheduled day due to unforeseen circumstances, but because it was the final session and the absence did not influence the results obtained in other sessions, the data were treated as missing at random. A total of 347 speech data sets from the test sessions (116 participants [4 groups] × 3 tests [pretest, posttest, delayed posttest] − 1) were included for analysis. An additional 696 speech data sets from the training sessions (58 participants [2 groups] × 12 performances [3 task repetitions × 4 sessions]) were transcribed in the same manner. The speech data were annotated using Praat (Boersma & Weenink, 2018), and silent pauses of 250 milliseconds or longer were identified with the automated detection of silent pauses on Praat and manually adjusted to ensure accuracy. Following Bui et al. (2019), the speech samples were co-coded by research assistants and one of the researchers. Specifically, the trained research assistants initially coded the speech performances using Praat, and the first author checked all the files to ensure accuracy. All disagreements were resolved through a discussion until consensus was reached. Consistent with previous studies on this topic (Bui et al., 2019; Y. Suzuki & Hanzawa, 2022), the fluency measures provided in Table 2 were used to cover speed, breakdown, and repair fluency (Skehan, 2003).

Statistical analysis

Analyses used generalized linear mixed-effects models (GLMMs), fitted with the lme4 package (version 1.1.27.1; Bates et al., 2015) in R (version 4.1.2; R Core Team, 2021). The significance of fixed effects was assessed with the Satterthwaite approximation for degrees of freedom using the lmerTest package (Kuznetsova et al., 2017). The probability distribution of the dependent variables was evaluated using density plots and Shapiro-Wilk tests. Due to positive skewness of fluency measures, the gamma distribution with the log link function was adopted. Following S. Suzuki (2021), the non-positive values (i.e., 0 values) were replaced by the −3 *SD* values of the theoretical distribution of the variable, estimated by the maximum likelihood estimation, to allow for the estimation of GLMMs based on a gamma distribution.

https://doi.org/10.1017/S0272263124000251 Published online by Cambridge University Press

For cases in which the dependent variable was normally distributed, linear mixed-effects models were used. All models included random intercepts of Participant and Task, as justified by the design. We initially considered the random slope of Time for Task (counterbalancing of the picture prompts). However, this maximal model (Barr et al., 2013) failed to converge; consequently, the random slope was removed.[3] The final code for the GLMMs is as follows:

*Fluency measure ~ Condition*Time + (1|Participant) + (1|Time)*

For effect size, Cohen's *d* was calculated based on an equation suggested by Westfall et al. (2014) for a design with random participants and random items[4]:

$$d = \frac{expected\ mean\ difference}{\sqrt{varintercept_{partipant} + varintercept_{task} + var_{residual}}}$$

In a meta-analysis of L2 distributed practice studies (S. K. Kim & Webb, 2022), the overall effect size of distributed practice effect (comparison between short- and long-spaced practice) was g = 0.40 for delayed posttest (a test was defined as a delayed posttest if it was administered as least 1 day after the treatment).[5] Hedge's *g* effect size is generally used for a small sample size (< 20), but the interpretation of its magnitude is identical to that of Cohen's *d*. Accordingly, the effect size above 0.40 was considered meaningful in the current study. This effect size is equivalent to small effect size according to a L2-general benchmark (Plonsky & Oswald, 2014): small (0.40), medium (0.70), and large (1.00).

Fluency Changes During Each Training Session

To investigate the effects of spacing schedules (1-day ISI vs. 7-day ISI) on training performance, we first sought to examine the group difference during each training session. To this end, we built a series of GLMMs for each training session: Training Session 2, Training Session 3, and Training Session 4 (note that Training Session 1 was not included as there was yet no difference in ISIs at the first training session). Because the first performance of each training session was assumed to be influenced by different spacing schedules, the fluency measure of the first performance (Time 2–1, Time 3–1, and Time 4–1) was treated as a covariate to control for the potential group difference at the beginning of each training session. The covariate variable was centered around its mean to reduce collinearity within the model (Cunnings, 2012). Each fluency measure was entered as an outcome variable, and Condition (1-day ISI, 7-day ISI), Time (performance 2, performance 3), Condition × Time interactions, and the covariate

[3]When a GLMM still failed to converge, we tried adjusting the optimizer (e.g., using "bobyqa"). For cases in which singularity fit warnings appeared, we inspected the results of the partially fit model and removed the random intercept that was accounting for a minimal amount of variance (i.e., random intercept of Task).

[4]The equation provided by Westfall et al. (2014) is for effect size calculation for a mixed-effects model with one fixed effect and two random factors. Given the more complex research design of the current study, the estimated effect sizes reported in this study should be interpreted with caution.

[5]As our study was conceptualized before the publication of S. K. Kim and Webb's (2022) paper, we could not conduct a priori power analysis using the effect sizes based on their findings. However, we used 0.40 as the effect size of interest based on their findings to interpret the results of the current study.

https://doi.org/10.1017/S0272263124000251 Published online by Cambridge University Press

were entered as the fixed-effect predictor variables. The fixed effects of Condition and Time were treatment-coded with the 1-day ISI condition at performance 2 as a reference level.

Fluency Changes Across Training Sessions

We also examined the fluency changes across multiple training sessions by narrowing down the scope of analysis to four specific time points (Time 1–1 [Training Session 1–performance 1], Time 2–1 [Training Session 2–performance 1], Time 3–1 [Training Session 3–performance 1], and Time 4–1 [Training Session 4–performance 1]). Time 1–1 served as a baseline, and each subsequent time point represented the critical difference in temporal spacing between the two conditions. Namely, at Time 2–1, Time 3–1, and Time 4–1, there was a lag of 1 day and 7 days from the previous training session for the 1-day ISI group and the 7-day ISI group, respectively. Each fluency measure was entered as a continuous outcome variable, and Condition, Time, and Condition × Time interactions were included as fixed-effect predictor variables. The fixed effects of Condition and Time were treatment coded with the 1-day ISI condition at Time 1–1 as a reference level.

Pretest–Posttest Changes

For the analysis of pretest–posttest changes, each fluency measure was entered as an outcome variable, and Condition (1-day ISI, control 1, 7-day ISI, control 2), Time (pretest, posttest, delayed posttest), and Condition × Time interactions were included as fixed-effect predictor variables. Rather than using treatment coding, the fixed effect of Condition was forward difference coded. In this coding scheme, the mean of the dependent variable for one categorical variable is compared with the mean of the dependent variable for the next, adjacent categorical variable. In the current analysis, the first contrast compared control group 1 with the 1-day ISI group (to investigate the effects of short-spaced fluency training); the second contrast compared the 1-day ISI group with the 7-day ISI group (to investigate the effects of different spacing conditions); and the third contrast compared the 7-day ISI group with control group 2 (to investigate the effects of long-spaced fluency training). In this way, additional multiple comparisons irrelevant to the predetermined objective of statistical analysis were avoided, thereby minimizing the risk of type I error. The fixed effect of Time was treatment-coded with the pretest as a reference level to identify any changes from the pretest to the posttests.

Results

Fluency changes during each training session

Figure 2 shows the performance scores during the training phase for each fluency measure (see Appendix A in Online Supplementary File for the descriptive statistics). Table 3 lists the effect sizes of group difference for each training session. In Training Session 2, the 7-day ISI group outperformed the 1-day ISI group in the third repetition in terms of articulation rate, mid-clause pause ratio, and repetition ratio, with effect sizes exceeding the benchmark of 0.40.

https://doi.org/10.1017/S0272263124000251 Published online by Cambridge University Press

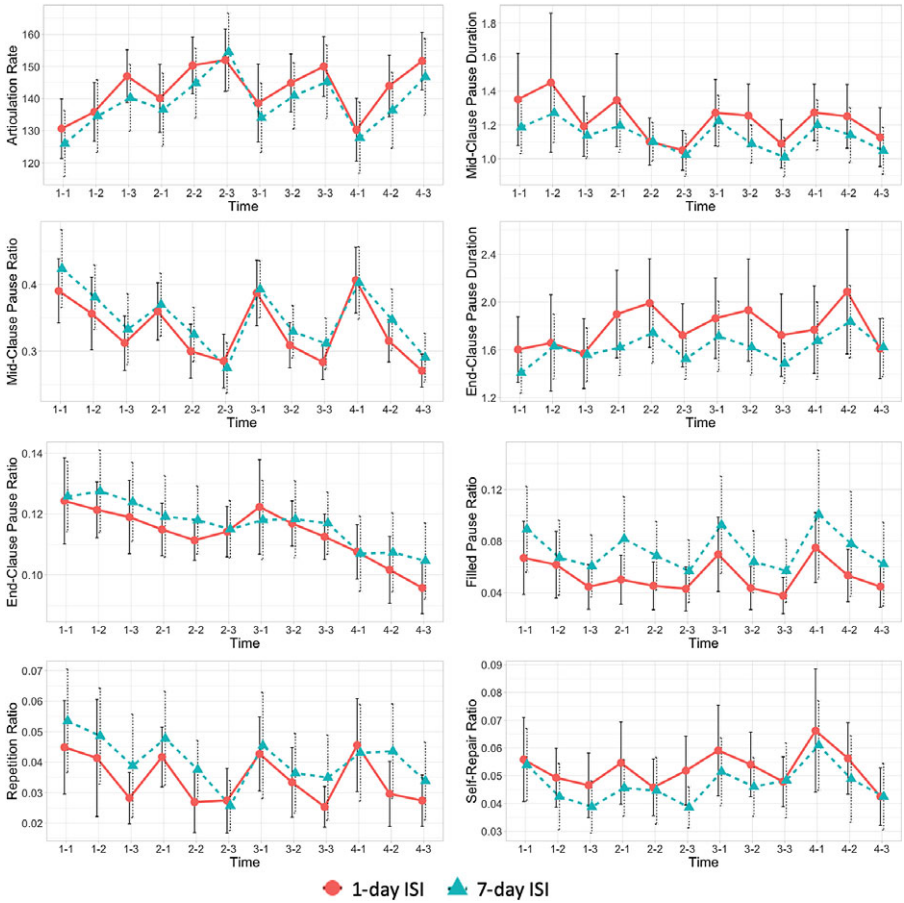

Figure 2. Performance scores during the training phrase.
Note: The error bars represent 95% confidence intervals.

Fluency changes across training sessions

In terms of fluency changes across multiple training sessions, the results indicated that there were no statistically significant Condition × Time interaction effects at any of the time points for any of the fluency measures (see Appendix E in the Online Supplementary File for model summaries). In other words, the participants assigned to the 1-day ISI and 7-day ISI conditions demonstrated similar performance at the beginning of each training session despite their difference in ISIs. Table 4 lists the effect sizes of group difference across training sessions.

Pretest–posttest changes

Figure 3 presents the mean test scores for the six fluency measures that showed statistically significant Condition × Time interaction effects. The descriptive statistics and detailed model summaries can be found in Appendices F–G in the Online

https://doi.org/10.1017/S0272263124000251 Published online by Cambridge University Press

https://doi.org/10.1017/S0272263124000251 Published online by Cambridge University Press

Table 3. Effect sizes (Cohen's d) for comparisons of 1-day ISI and 7-day ISI conditions during each training session

	Training Session 2		Training Session 3		Training Session 4	
Fluency measure	Time 2–2	Time 2–3	Time 3–2	Time 3–3	Time 4–2	Time 4–3
Articulation rate	−0.17 [−0.69, 0.35]	0.49* [0.06, 0.93]	−0.03 [9.97, 10.71]	−0.06 [−0.50, 0.39]	−0.37 [−0.88, 0.15]	0.19 [−0.27, 0.64]
Mid-clause pause duration	0.30 [−0.40, 1.00]	−0.22 [−0.59, 0.15]	−0.46 [−1.14, 0.21]	0.24 [−0.16, 0.63]	−0.16 [−1.00, 0.68]	0.20 [−0.32, 0.72]
Mid-clause pause ratio	0.20 [−0.51, 0.91]	−0.57* [−0.93, −0.21]	0.15 [−0.53, 0.84]	0.12 [−0.27, 0.51]	0.36 [−0.34, 1.06]	−0.15 [−0.52, 0.23]
End-clause pause duration	0.00 [−0.62, 0.63]	0.02 [−0.43, 0.47]	−0.17 [−0.89, 0.54]	0.12 [−0.21, 0.45]	−0.06 [−1.12, 1.00]	0.24 [−0.42, 0.90]
End-clause pause ratio	0.12 [−0.51, 0.75]	−0.23 [−0.70, 0.24]	0.02 [−0.62, 0.66]	0.23 [−0.24, 0.70]	0.16 [−0.2, 0.52]	0.08 [−0.22, 0.38]
Filled pause ratio	−0.09 [−0.75, 0.57]	−0.33 [−0.67, 0.00]	−0.10 [−0.75, 0.55]	0.02 [−0.36, 0.39]	−0.06 [−0.74, 0.63]	−0.25 [−0.51, 0.01]
Repetition ratio	0.56 [−0.08, 1.19]	−0.58* [−1.04, −0.13]	0.05 [−0.58, 0.69]	0.11 [−0.34, 0.57]	0.46 [−0.18, 1.09]	−0.19 [−0.59, 0.21]
Self-repair ratio	−0.03 [−0.67, 0.61]	−0.19 [−0.70, 0.33]	−0.19 [−0.83, 0.45]	0.08 [−0.63, 0.80]	−0.45 [−1.10, 0.19]	0.30 [−0.25, 0.86]

Note: A positive effect size indicates a higher value for the 7-day ISI group, whereas a negative effect size indicates a higher value for the 1-day ISI group. The values in brackets indicate 95% confidence intervals. See Appendices B–D in the Online Supplementary File for the full results.

*$p < .05$.

Table 4. Effect sizes (Cohen's d) for comparisons of 1-day ISI and 7-day ISI conditions across training sessions

Fluency measure	Time 2–1	Time 3–1	Time 4–1
Articulation rate	0.04 [−0.30, 0.39]	0.00 [−0.34, 0.35]	0.08 [−0.27, 0.42]
Mid-clause pause duration	0.02 [−0.40, 0.43]	0.20 [−0.21, 0.62]	0.10 [−0.31, 0.52]
Mid-clause pause ratio	−0.19 [−0.61, 0.22]	−0.17 [−0.58, 0.25]	−0.35 [−0.77, 0.07]
End-clause pause duration	−0.05 [−0.45, 0.34]	0.06 [−0.34, 0.45]	0.17 [−0.22, 0.57]
End-clause pause ratio	0.01 [−0.51, 0.53]	−0.19 [−0.71, 0.33]	−0.15 [−0.67, 0.36]
Filled pause ratio	−0.13 [−0.53, 0.27]	−0.01 [−0.41, 0.39]	−0.10 [−0.50, 0.30]
Repetition ratio	−0.06 [−0.61, 0.50]	−0.12 [−0.66, 0.43]	−0.37 [−0.92, 0.18]
Self-repair ratio	−0.10 [−0.69, 0.50]	−0.15 [−0.74, 0.45]	0.04 [−0.55, 0.63]

Note. A positive effect size indicates a higher value for the 7-day ISI group, whereas a negative effect size indicates a higher value for the 1-day ISI group. The values in brackets indicate 95% confidence intervals. See Appendix E in the Online Supplementary File for the full results.

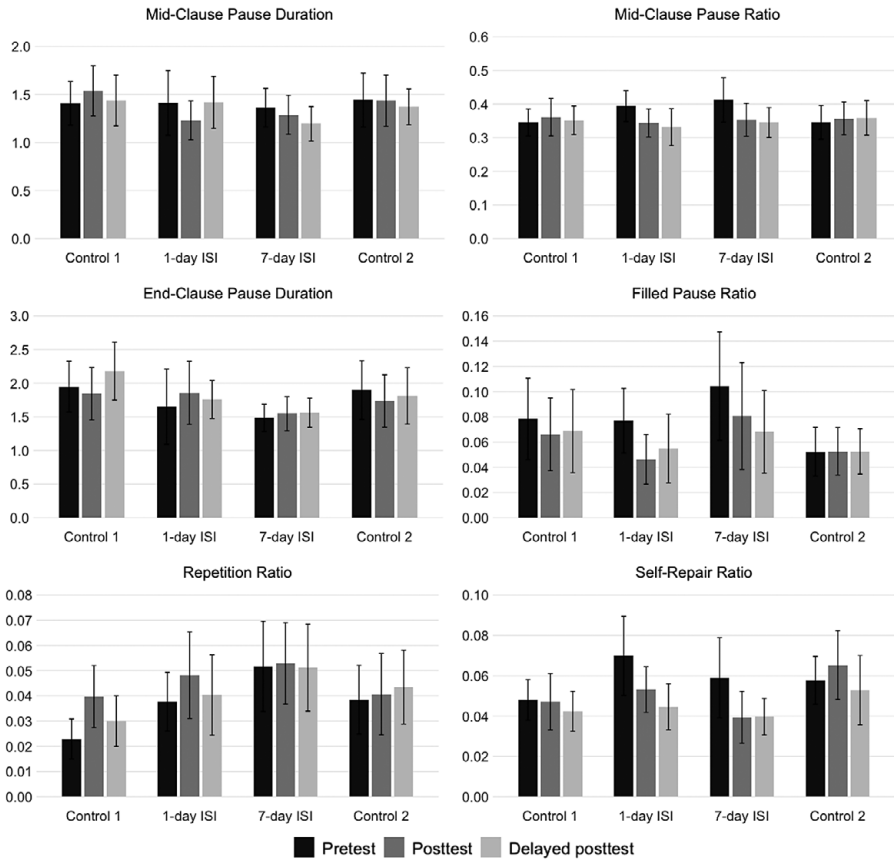

Figure 3. Test scores for the four groups.
Note: The error bars represent 95% confidence intervals.

https://doi.org/10.1017/S0272263124000251 Published online by Cambridge University Press

Supplementary File.[6] In what follows, we discuss the results for each aspect of fluency (speed, breakdown, and repair).

Speed fluency

Table 5 summarizes the GLMM results for the pretest–posttest changes. In terms of speed fluency (articulation rate), the differences in gains between the experimental groups (the 1-day ISI and 7-day ISI groups) and the corresponding control groups as well as between the 1-day ISI and 7-day ISI groups were not statistically significant at either posttest or delayed posttest. The results thus indicate that fluency practice or its distribution had little impact on the development of speed fluency.

Breakdown Fluency

The 1-day ISI group decreased the mid-clause pause duration in the first posttest relative to the corresponding control group with a meaningful effect size (p = .015, d = 0.48, 95% CI [0.09, 0.87]). The 1-day ISI group also reduced the mid-clause pause ratio relative to the corresponding control group with effect sizes above the benchmark in the first posttest (p = .016, d = 0.49, 95% CI [0.05, 0.92]) and delayed posttest (p = .003, d = 0.61, 95% CI [0.17, 1.05]). Likewise, the 7-day ISI group reduced the mid-clause pause ratio relative to the corresponding control group with effect sizes above the benchmark in the first posttest (p = .007, d = −0.54, 95% CI [−0.97, −0.11]) and delayed posttest (p = .003, d = −0.58, 95% CI [−1.01, −0.15]). These findings suggest that fluency training was effective in reducing the number of mid-clause pauses regardless of ISIs.

The results for end-clause pause duration showed a fluency change in an unexpected direction: The 1-day ISI group increased end-clause pause duration in the first posttest relative to the corresponding control group with a meaningful effect size (p = .036, d = −0.40, 95% CI [−0.77, −0.03]). For filled pause ratio, the 7-day ISI group made a greater improvement compared with the corresponding control group with an effect size slightly under the benchmark in the first posttest (p = .048, d = −0.36, 95% CI [−0.72, 0.00]) and well above the benchmark in the delayed posttest (p = .004, d = −0.52, 95% CI [−0.87, −0.16]). The 1-day ISI and 7-day ISI groups did not show any statistically significant differences in gains in any of the breakdown fluency measures.

Repair Fluency

The difference in gains between the 1-day ISI group and the corresponding control group for repetition ratio was marginally significant in the first posttest (p = .059, d = 0.50, 95% CI [−0.02, 1.03]) and statistically significant in the delayed posttest (p = .047, d = 0.55, 95% CI [0.01, 1.09]), both with effect sizes above the benchmark. As shown in Figure 3, these two groups in fact increased the number of repetitions from the pretest to each posttest; however, the positive effect sizes suggest that the degree of

[6]GLMMs showed that there was no significant group difference at pretest for any fluency measures, with the exception of end-clause pause ratio between the 1-day ISI group and the corresponding control group (p = .03). An additional GLMM was built for this measure by entering the pretest score as a covariate, along with the fixed-effects variables of Condition, Time (posttest, delayed posttest), and Condition × Time interactions. The results showed no significant simple effects or interaction effects at either posttest (p > .05).

https://doi.org/10.1017/S0272263124000251 Published online by Cambridge University Press

Table 5. Summary of the GLMM results for Condition × Time interactions

		Posttest			Delayed posttest		
Fluency measure	Contrast	*z/t*	*p*	*d*	*z/t*	*p*	*d*
Articulation rate	Control 1 vs. 1–day ISI	−0.304	.762	−0.06 [−0.42, 0.30]	0.343	.732	0.06 [−0.30, 0.43]
	1–day ISI vs. 7–day ISI	−0.307	.759	−0.06 [−0.41, 0.30]	−0.078	.938	−0.01 [−0.37, 0.35]
	7–day ISI vs. Control 2	0.556	.579	0.10 [−0.25, 0.46]	0.381	.704	0.07 [−0.29, 0.42]
Mid-clause pause duration	Control 1 vs. 1–day ISI	2.435	.015	0.48* [0.09, 0.87]	−0.021	.983	0.00 [−0.40, 0.39]
	1–day ISI vs. 7–day ISI	−0.423	.673	−0.08 [−0.47, 0.30]	1.878	.060	0.37 [−0.02, 0.76]
	7–day ISI vs. Control 2	−0.866	.387	−0.17 [−0.55, 0.21]	−1.626	.104	−0.32 [−0.70, 0.07]
Mid-clause pause ratio	Control 1 vs. 1–day ISI	2.405	.016	0.49* [0.05, 0.92]	2.988	.003	0.61* [0.17, 1.05]
	1–day ISI vs. 7–day ISI	0.204	.838	0.04 [−0.40, 0.48]	−0.374	.709	−0.08 [−0.51, 0.36]
	7–day ISI vs. Control 2	−2.719	.007	−0.54* [−0.97, −0.11]	−2.938	.003	−0.58* [−1.01, −0.15]
End-clause pause duration	Control 1 vs. 1–day ISI	−2.101	.036	−0.40* [−0.77, −0.03]	−0.188	.851	−0.04 [−0.41, 0.34]
	1–day ISI vs. 7–day ISI	1.093	.274	0.21 [−0.16, 0.57]	0.972	.331	0.18 [−0.19, 0.56]
	7–day ISI vs. Control 2	1.333	.183	0.25 [−0.12, 0.61]	0.999	.318	0.19 [−0.18, 0.55]
End-clause pause ratio	Control 1 vs. 1–day ISI	−1.121	.262	−0.29 [−0.80, 0.22]	−1.760	.078	−0.46 [−0.97, 0.05]
	1–day ISI vs. 7–day ISI	0.402	.687	0.10 [−0.40, 0.61]	1.804	.071	0.47 [−0.04, 0.97]
	7–day ISI vs. Control 2	−0.424	.672	−0.11 [−0.61, 0.39]	−0.235	.814	−0.06 [−0.56, 0.44]
Filled pause ratio	Control 1 vs. 1–day ISI	1.402	.161	0.25 [−0.10, 0.61]	0.696	.486	0.13 [−0.23, 0.48]
	1–day ISI vs. 7–day ISI	−0.736	.462	−0.13 [−0.49, 0.22]	0.755	.451	0.14 [−0.22, 0.49]
	7–day ISI vs. Control 2	−1.977	.048	−0.36* [−0.72, 0.00]	−2.860	.004	−0.52* [−0.87, −0.16]
Repetition ratio	Control 1 vs. 1–day ISI	1.888	.059	0.50 [−0.02, 1.03]	1.983	.047	0.55* [0.01, 1.09]
	1–day ISI vs. 7–day ISI	−0.018	.986	0.00 [−0.52, 0.51]	−0.970	.332	−0.26 [−0.79, 0.27]
	7–day ISI vs. Control 2	0.637	.524	0.16 [−0.34, 0.67]	−0.101	.920	−0.03 [−0.53, 0.48]
Self-repair ratio	Control 1 vs. 1–day ISI	0.378	.705	0.12 [−0.50, 0.73]	1.248	.212	0.39 [−0.22, 1.01]
	1–day ISI vs. 7–day ISI	1.008	.313	0.31 [−0.29, 0.92]	−0.723	.470	−0.22 [−0.83, 0.38]
	7–day ISI vs. Control 2	−2.253	.024	−0.68* [−1.28, −0.09]	−0.249	.804	−0.08 [−0.68, 0.53]

Note: The values in brackets indicate 95% confidence intervals. See Appendix G in the Online Supplementary File for the full results.
*$p < .05$.

https://doi.org/10.1017/S0272263124000251 Published online by Cambridge University Press

increase was smaller for the 1-day ISI group than for the control group. Finally, the 7-day ISI group reduced the self-repair ratio relative to the corresponding control group with a meaningful effect size in the first posttest (p = .024, d = −0.68, 95% CI [−1.28, −0.09]). The difference in gains between the 1-day ISI and 7-day ISI groups was not statistically significant for repair fluency measures.

Discussion

Fluency changes during each training session

The analyses of fluency changes during Training Session 2 showed that the 7-day ISI group outperformed the 1-day ISI group in the third repetition. Specifically, the 7-day ISI group improved articulation rate, mid-clause pause ratio, and repetition ratio relative to the 1-day ISI group with meaningful effect sizes (d = |0.49–0.58|). As shown in Figure 2, the developmental patterns for the two groups are similar until the second repetition, but the 7-day ISI group demonstrated greater further gains in the third repetition than did the 1-day ISI group. This is contrary to our hypothesis, as we had expected superior training performance from the 1-day ISI group due to the reduced cognitive load for linguistic formulation yielded by shorter ISIs. Although shorter spacing has presumably helped the learners to better remember how to perform the narrative task using their linguistic repertoires, the results suggest that some speakers in the 1-day ISI condition reached a plateau in the second repetition and had little room for further improvement in the third repetition. Alternatively, the reduced cognitive demands in the 1-day ISI condition may have shifted the speakers' attention to other aspects of speaking performance such as syntactic complexity or accuracy. In other words, some speakers might have aimed for a better performance with more complex and/or accurate grammar and vocabulary, which led to a decline in fluency gains in the third repetition. Longer spacing, by contrast, has presumably induced greater difficulty in the task repetition practice; however, the additional difficulty induced by longer ISIs might have ensured that speakers gradually improved their fluency, challenging them to expend greater effort and make further improvements in the final repetition.

It is worthy of note that statistically significant group differences were only observed in Training Session 2. This might be due to the possibility that the effects of spacing intervals diminished as the number of training sessions increased. In other words, the impact of practice distribution appeared larger in the beginning of training and then subsided in the later training sessions (see Bahrick et al., 1993).

Fluency changes across training sessions

The analyses of fluency changes across multiple training sessions showed that despite the difference in ISIs, the 1-day ISI and 7-day ISI groups demonstrated similar performance at each critical time point (i.e., Time 2–1, Time 3–1, Time 4–1). The lack of statistical significance might be ascribed to the fact that a new picture prompt was used in each training session. As discussed in the literature review section, when speakers repeat the same task, fluency improves on the repeated performance due to the reduced cognitive load for conceptualization and formulation (Lambert et al., 2017). However, when different tasks are used (i.e., procedural repetition), fluency does not improve as much because speakers have to engage in the processes of conceptualization and formulation afresh (Lambert et al., 2021). The current findings suggest that the participants in both groups faced similar cognitive demands at the

https://doi.org/10.1017/S0272263124000251 Published online by Cambridge University Press

beginning of each training session, as they equally had to engage in a novel task. It is speculated that greater group differences might be observed if an identical task is used across multiple training sessions (Y. Suzuki & Hanzawa, 2022). It is possible, however, that there could be diminishing returns if the same tasks were repeated across multiple training sessions, as it might lead to boredom at some point (Hanzawa & Suzuki, 2023; Lambert et al., 2017).

An important question to be answered is what is repeated in procedural repetition. Although the participants in the current study were not expected to retrieve the specific linguistic items used to narrate a certain story (e.g., *vase*, *airport*), there was some overlap in linguistic content between different tasks due to the similarities in the task design features (e.g., chronological sequence, characters, an element of surprise in the narrative). As such, it is likely that learners repeated linguistic expressions that were useful across different tasks (e.g., *and then*, *children*, *surprisingly*). In addition, the nature of the picture-description tasks requires the learners to rely on the (re)use of the same syntactic structures such as verb tenses (e.g., past tense) and part-of-speech trigrams (e.g., *the taller boy*; determiner–adjective–noun) while engaging in tasks of the same type, which aids proceduralization of L2 speech processing (N. de Jong & Tillman, 2018; Y. Suzuki et al., 2022). The framework of desirable difficulties embodies a variety of instructional practices that induce optimal difficulty in the learning process (e.g., distributing practice, interleaving materials, providing contextual interference, reducing feedback) (Bjork, 1994). Varying the instructional tasks might represent a desirably difficult pedagogical approach for L2 speaking skills development, as it familiarizes the learners with the general procedural demands of the given task types and develops their flexibility in applying linguistic knowledge and skills to new task contexts (DeKeyser, 2018; Lambert et al., 2021; Larsen-Freeman, 2018).

Effects of spacing intervals on pretest–posttest changes

Our second research question examined the effects of spacing intervals on fluency development as demonstrated by pretest–posttest–delayed posttest changes. The findings showed no significant differences between the two spaced conditions, indicating that short-spaced practice and long-spaced practice were equally beneficial for improving L2 fluency overall. These findings are largely consistent with the broader body of distributed practice research in cognitive psychology (e.g., Cepeda et al., 2009) and SLA (e.g., Nakata, 2015), which have shown minimal effects when comparing two or more spaced conditions (lag effects), as opposed to large effects when comparing a massed condition against spaced conditions (spacing effects) (see Rogers, 2023). The current findings are also consistent with Y. Suzuki and Hanzawa's (2022) study on distributed L2 fluency practice, which revealed no significant difference between short- and long-spaced conditions.

The lack of significant differences might be attributed to several methodological factors. First, the learners in both experimental groups engaged in massed task repetition within each training session (i.e., three successive narrative performances using the same task). Although repeating the same task in immediate succession arguably increased the impact of fluency training (N. de Jong & Perfetti, 2011; Y. Suzuki, 2021), it might have reduced the effects of practice distribution, leading to similar fluency gains for the short- and long-spaced conditions. The results of training performance indicated that the benefits of longer-spaced practice were best realized in the third task repetition. Thus, a different pattern of learning gains might have been

https://doi.org/10.1017/S0272263124000251 Published online by Cambridge University Press

observed if each training session involved only one or two task performances (instead of three as in the current study).

Another methodological issue is that RIs were manipulated *within* participants rather than *between* participants. In other words, the same participants took the posttest and the delayed posttest after completing the training sessions. The posttests were administered using novel prompts, just as in the training phase; therefore, the first posttest can be considered another practice opportunity, which may have influenced the performance in the delayed posttest (Y. Suzuki, 2017). If the first posttest was treated as another training session, the average ISI would increase to 2.5 days ([1+1+1 +7]/4) for the short-spaced condition, and no change would be made for the long-spaced condition (7 days). For the recalculated RI of 21 days, the ratio of ISI to the delayed posttest would be 12% (originally 4%) and 33% (originally 25%) for the short- and long-spaced conditions, respectively. These recalculated values are approximately within the optimal range (10%–30%), which might explain why the two groups demonstrated similar performances in the delayed posttest.

Still another explanation for the non-significant findings is the possibility that the ISI–RI ratios used in the current study were suboptimal. According to Cepeda et al. (2008), the ideal proportion of an ISI–RI ratio depends on the length of a given RI. For instance, for the RI of 35 days, their study showed that the optimal ISI–RI ratio was 23% for the recall test and 19% for the recognition test. In contrast, for a much shorter RI of 7 days, the optimal ratios were 43% and 24% for the recall and recognition tests, respectively. In the current study, the ratios of the ISI to the first posttest (7-day RI) were 14% and 100% for the short- and long-spaced conditions, respectively. These values both depart from the optimal value suggested for a 7-day RI posttest (i.e., 24%–43%); consequently, the two spaced conditions might have led to similar fluency gains at first posttest in the current study. Rather than using the range of 10%–30% as a general rule of thumb, it might be worthwhile to use optimal ISI–RI ratios directly in light of the findings from Cepeda et al. (2008).

The results showed that, regardless of ISIs, the effects of fluency training were pronounced in terms of mid-clause pause ratio (d = |0.49–0.61|). Pauses within the clause boundaries are particularly related to L2 linguistic encoding processes (N. H. de Jong, 2016; Kahng, 2018); thus, the findings suggest that the participants have proceduralized some aspects of the linguistic formulation processes that underlie L2 fluency (e.g., lexical retrieval, syntactic construction). The current findings are in line with previous research on fluency training (N. de Jong & Perfetti, 2011) and further demonstrate that oral practice using a combination of same-task repetition and procedural repetition is a useful strategy for facilitating L2 fluency transfer.

Other fluency measures indicated differential effects of spacing on fluency development. The 7-day ISI group showed a marked improvement in filled pause ratio, indicating that the benefit of longer-spaced practice extends to the reduction of both silent and filled pauses. As for repair fluency, the 7-day ISI group demonstrated decreased self-repair ratio in the first posttest relative to the corresponding control group. This finding may be partially in line with previous research (Bui et al., 2019; Y. Suzuki & Hanzawa, 2022) that showed the benefit of longer spacing for repair fluency (i.e., reducing self-repetitions). From a theoretical perspective, however, self-repairs and self-repetitions play different roles in speech production, with the former reflecting one's self-monitoring endeavors (Kormos, 2006) rather than breakdown in language processing. Thus, the findings obtained in the current study may extend those reported by previous studies in that they indicate the impact of longer spacing on L2 speakers' self-monitoring processes.

https://doi.org/10.1017/S0272263124000251 Published online by Cambridge University Press

The benefits of shorter spacing were also observed on some aspects of fluency. Particularly, practice under the 1-day ISI condition led to a significant improvement in mid-clause pause duration in the first posttest. Although the difference between the 1-day ISI and 7-day ISI conditions was not statistically significant, the descriptive statistics showed that the mean changes from the pretest to the first posttest were –0.18 and –0.07 for the 1-day ISI group and the 7-day ISI group, respectively, indicating a greater improvement for the shorter-spaced condition. The 1-day ISI group, however, exhibited longer end-clause pause duration relative to the corresponding control group. Thus, the improvement observed in mid-clause pause duration may be simply reflecting the trade-off effects (i.e., mid-clause pause duration decreasing at the cost of end-clause pause duration). However, as pausing behaviors at the clausal boundary are related to conceptual planning of the speech content (N. H. de Jong, 2016; Kahng, 2018), longer end-clause pauses may not necessarily indicate decreased performance. Rather, the speakers in the 1-day ISI group may have learned to pause more effectively at the appropriate clausal boundary (Y. Suzuki, 2021).

Both the 1-day ISI group and the corresponding control group increased the number of verbatim repetitions from the pretest to the posttests. As discussed in the Results section, the extent of increase was smaller for the 1-day ISI group than for the corresponding control group, indicating that fluency practice may have attenuated the rise in repetition ratio. The other two groups with longer pretest–posttest intervals (7-day ISI group and control group 2) did not show such a rising pattern in repetition ratio. Thus, these findings suggest that longer spacing might be generally more beneficial than shorter spacing for maintaining or improving repetition frequency (Bui et al., 2019; Y. Suzuki & Hanzawa, 2022).

While short- and long-spaced practice led to similar fluency gains overall, the two groups demonstrated different developmental trajectories in terms of mid-clause pause duration. As Figure 3 shows, the developmental pattern of the 1-day ISI group is characterized by an initial improvement at posttest, followed by a decay at delayed posttest, whereas the developmental pattern of the 7-day ISI group shows a gradual and steady improvement overall. These observations were supported by within-group comparisons: The difference between the pretest and the posttest was statistically significant for the 1-day ISI group (p = .050, d = −0.28, 95% CI [−0.56, 0.00]), whereas the difference between the pretest and the delayed posttest was statistically significant for the 7-day ISI group (p = .011, d = −0.35, 95% CI [−0.62, −0.08]). The corresponding control groups, by contrast, did not reveal any significant within-group changes at either posttest or delayed posttest (p > .05). These patterns suggest that the 7-day ISI group may have retained the effects of fluency training for a longer period than did the 1-day ISI group, lending support to previous research that has demonstrated more durable learning gains for longer-spaced practice (e.g., Bird, 2010; Cepeda et al., 2008). Given that these developmental patterns were observed for only one fluency measure (mid-clause pause duration), further research is required to confirm whether these observations hold true in other study contexts.

Limitations and directions for future research

The current study has several limitations that need to be addressed in future research. First, as noted earlier, the current methodological design that involved massed task repetitions during each training session makes it difficult to interpret the effects of ISIs in isolation from the effects of task repetition. Future research should therefore investigate the effects of ISIs by partialling out the effects of massed task repetitions.

https://doi.org/10.1017/S0272263124000251 Published online by Cambridge University Press

Second, the effects of ISIs were potentially confounded with the effects of RIs in the current research design that used RIs as a within-participants factor. Thus, it would be worth adopting a study design that uses multiple RIs as a between-participants factor in future research (Muqaibal et al., 2023).

Third, the effects of distributed practice were examined using only monologue narrative tasks. Although this type of task is useful for controlling different variables (e.g., interlocutor influence), it does not represent the interactive communication skills that L2 speakers need in authentic contexts. Moreover, picture-narrative tasks are closed tasks, in which the content of speech is predefined by the given prompt (Pallotti, 2009). Thus, it is speculated that the speakers' attentional resources were primarily directed to linguistic formulation rather than conceptualization (Levelt, 1989). In future research, open tasks (e.g., opinion tasks) should be used to elucidate how task type may moderate distributed practice effects in fluency development. Finally, there could a host of individual difference factors that may have moderated the effects of distributed practice (e.g., aptitude, proficiency, task motivation). As learner-related characteristics play an important role in making L2 practice desirably difficult (Serfaty & Serrano, 2022; Y. Suzuki et al., 2019), future researchers should consider the impact of these individual difference factors in speaking skills development.

Conclusion

The objective of the current study was to investigate the effects of distributed practice on L2 fluency development by applying the optimal ISI–RI ratio informed by cognitive psychology research (Cepeda et al., 2008; Rohrer & Pashler, 2007). The findings indicated that while short- and long-spaced practice led to different developmental trajectories during the training phase, the two groups demonstrated similar fluency gains overall in the two posttests. The current study extends previous distributed practice research on L2 speaking skills development (Bui et al., 2019; Kobayashi, 2022; Y. Suzuki & Hanzawa, 2022) and adds to the larger body of knowledge in cognitive psychology (e.g., Cepeda et al., 2008) and SLA (e.g., Kasprowicz et al., 2019) investigating the role of ISI–RI ratios in distributed practice. As this study represents the first attempt to explore the effects of distributed practice on L2 fluency development through the manipulation of ISI–RI ratios, more comprehensive future studies are needed to draw definitive conclusions.

Supplementary material. The supplementary material for this article can be found at http://doi.org/10.1017/S0272263124000251.

Acknowledgments. We thank the editor and the anonymous reviewers for their insightful comments and suggestions on the earlier versions of our paper. We also thank Shungo Suzuki and Toshitaka Hamamura for sharing their expertise in data coding and statistical analysis. We also acknowledge Joshua Kidd for his valuable feedback on the earlier drafts of this article. This study was supported by JSPS KAKENHI (Grant No. 20K13102).

Data availability statement. The data and materials are available at https://iris-database.org.

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*(5), 316–321. https://doi.org/10.1111/j.1467-9280.1993.tb00571.x

https://doi.org/10.1017/S0272263124000251 Published online by Cambridge University Press

Bahrick, H. P., & Hall, L. K. (2005). The importance of retrieval failures to long-term retention: A metacognitive explanation of the spacing effect. *Journal of Memory and Language*, *52*(4), 566–577. https://doi.org/10.1016/j.jml.2005.01.012

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bird, S. (2010). Effects of distributed practice on the acquisition of second language English syntax. *Applied Psycholinguistics*, *31*(4), 635–650. https://doi.org/10.1017/S0142716410000172

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). MIT Press.

Boersma, P., & Weenink, D. (2018). *Praat: Doing phonetics by computer [Computer program]*. http://www.praat.org

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

Bygate, M. (2018). *Learning language through task repetition*. John Benjamins.

Cepeda, N. J., Coburn, N., Rohrer, D., Wixted, J. T., Mozer, M. C., & Pashler, H. (2009). Optimizing distributed practice: Theoretical analysis and practical implications. *Experimental Psychology*, *56*(4), 236–246. https://doi.org/10.1027/1618-3169.56.4.236

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

Cepeda, N. J., Vul, E., Rohrer, D., Wixted, J., & Pashler, H. (2008). Spacing effects in learning: A temporal ridgeline of optimal retention. *Psychological Science*, *19*(11), 1095–1102. https://doi.org/10.1111/j.1467-9280.2008.02209.x

Cunnings, I. (2012). An overview of mixed-effects statistical models for second language researchers. *Second Language Research*, *28*(3), 369–382. https://doi.org/10.1177/0267658312443651

de Jong, N. H. (2016). Predicting pauses in L1 and L2 speech: The effects of utterance boundaries and word frequency. *International Review of Applied Linguistics in Language Teaching*, *54*(2), 113–132. https://doi.org/10.1515/iral-2016-9993

de Jong, N., & Perfetti, C. A. (2011). Fluency training in the ESL classroom: An experimental study of fluency development and proceduralization. *Language Learning*, *61*(2), 533–568. https://doi.org/10.1111/j.1467-9922.2010.00620.x

de Jong, N., & Tillman, P. (2018). Grammatical structures and oral fluency in immediate task repetition. In M. Bygate (Ed.), *Learning language through task repetition* (pp. 43–73). John Benjamins.

de Jong, N., & Vercellotti, M. L. (2016). Similar prompts may not be similar in the performance they elicit: Examining fluency, complexity, accuracy, and lexis in narratives from five picture prompts. *Language Teaching Research*, *20*(3), 387–404. https://doi.org/10.1177/1362168815606161

DeKeyser, R. (2015). Skill acquisition theory. In J. VanPatten, B. & Williams (Ed.), *Theories in second language acquisition: An introduction* (pp. 94–112). Routledge.

DeKeyser, R. (2018). Task repetition for language learning: A perspective from skill acquisition theory. In M. Bygate (Ed.), *Learning language through task repetition* (pp. 27–41). John Benjamins.

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. https://doi.org/10.1037/0021-9010.84.5.795

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, Supplement, *14*(1), 4–58. https://doi.org/10.1177/1529100612453266

Ebbinghaus, H. (1885). *Memory: A contribution to experimental psychology (H. A. Ruger, C. E. Bussenius, & E. R. Hilgard, Trans.)*. Dover Publications.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191.

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics*, *21*(3), 354–375. https://doi.org/10.1093/applin/21.3.354

https://doi.org/10.1017/S0272263124000251 Published online by Cambridge University Press

Hanzawa, K., & Suzuki, Y. (2023). How do learners perceive task repetition? Distributed practice effects on engagement and metacognitive judgment. *The Modern Language Journal*, 1–28. https://doi.org/10.1111/modl.12843

Heaton, J. B. (1966). *Composition through pictures*. Longman.

Heaton, J. B. (1975). *Beginning composition through pictures*. Longman.

Kahng, J. (2018). The effect of pause location on perceived fluency. *Applied Psycholinguistics*, *39*(3), 569–591. https://doi.org/10.1017/S0142716417000534

Kakitani, J. (2023). Equivalency of picture-based speaking tasks: An investigation of complexity, accuracy, lexis, and fluency. *The Language Teacher*, *47*(2), 3–10. https://doi.org/10.37546/JALTTLT47.2-1

Kasprowicz, R. E., Marsden, E., & Sephton, N. (2019). Investigating distribution of practice effects for the learning of foreign language verb morphology in the young learner classroom. *The Modern Language Journal*. https://doi.org/10.1111/modl.12586

Kim, A. S. N., Wong-Kee-You, A. M. B., Wiseheart, M., & Rosenbaum, R. S. (2019). The spacing effect stands up to big data. *Behavior Research Methods*, *51*(4), 1485–1497. https://doi.org/10.3758/s13428-018-1184-7

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. https://doi.org/10.1080/1464536X.2011.573008

Kim, S. K., & Webb, S. (2022). The effects of spaced practice on second language learning: A meta-analysis. *Language Learning*, *72*(1), 269–319. https://doi.org/10.1111/lang.12479

Kim, Y., & Tracy-Ventura, N. (2013). The role of task repetition in L2 performance development: What needs to be repeated during task-based interaction? *System*, *41*(3), 829–840. https://doi.org/10.1016/j.system.2013.08.005

Kobayashi, M. (2022). The distributed practice effects of speaking task repetition. *International Journal of Applied Linguistics*, *32*(1), 142–157. https://doi.org/10.1111/ijal.12409

Kormos, J. (2006). *Speech production and second language acquisition*. Routledge.

Koval, N. G. (2019). Testing the deficient processing account of the spacing effect in second language vocabulary learning: Evidence from eye tracking. *Applied Psycholinguistics*, *40*(5), 1103–1139. https://doi.org/10.1017/S0142716419000158

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lambert, C., Aubrey, S., & Leeming, P. (2021). Task preparation and second language speech production. *TESOL Quarterly*, *55*(2), 331–365. https://doi.org/10.1002/tesq.598

Lambert, C., Kormos, J., & Minn, D. (2017). Task repetition and second language speech processing. *Studies in Second Language Acquisition*, *39*(1), 167–196. https://doi.org/10.1017/S0272263116000085

Larsen-Freeman, D. (2018). Task repetition or task iteration?: It does make a difference. In M. Bygate (Ed.), *Learning language through task repetition* (pp. 311–329). John Benjamins.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. MIT Press.

Miles, S. W. (2014). Spaced vs. Massed distribution instruction for L2 grammar learning. *System*, *42*(1), 412–428. https://doi.org/10.1016/j.system.2014.01.014

Muqaibal, M. H., Kasprowicz, R., & Tissot, C. (2023). Evaluating the impact of spaced practice using computer-assisted language learning (CALL) on vocabulary learning in the classroom. *Language Teaching Research*, 1–38. https://doi.org/10.1177/13621688221146146

Murayama, K., Usami, S., & Sakaki, M. (2022). Summary-statistics-based power analysis: A new and practical method to determine sample size for mixed-effects modeling. *Psychological Methods*, *27*(6), 1014–1038. https://doi.org/10.1037/met0000330

Nakata, T. (2015). Effects of expanding and equal spacing on second language vocabulary learning: Does gradually increasing spacing increase vocabulary learning. *Studies in Second Language Acquisition*, *37*(4), 677–711. https://doi.org/10.1017/S0272263114000825

Pallotti, G. (2009). CAF: Defining, refining and differentiating constructs. *Applied Linguistics*, *30*(4), 590–601. https://doi.org/10.1093/applin/amp045

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

R Core Team. (2021). *R: A language and environment for statistical computing*. https://www.r-project.org/

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

https://doi.org/10.1017/S0272263124000251 Published online by Cambridge University Press

Rogers, J. (2023). Spacing effects in task repetition research. *Language Learning*, *73*(2), 445–474. https://doi.org/10.1111/lang.12526

Rohrer, D. (2015). Student instruction should be distributed over long time periods. *Educational Psychology Review*, *27*(4), 635–643. https://doi.org/10.1007/s10648-015-9332-4

Rohrer, D., & Pashler, H. (2007). Increasing retention without increasing study time. *Psychological Science*, *16* (4), 183–186. https://doi.org/10.1111/j.1467-8721.2007.00500.x

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

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

Serrano, R. (2022). A state-of-the-art review of distribution-of-practice effects on L2 learning. *Studies in Second Language Learning and Teaching*, *12*(3), 355–379. https://doi.org/10.14746/ssllt.2022.12.3.2

Skehan, P. (2003). Task-based instruction. *Language Teaching*, *36*, 1–14. https://doi.org/10.1017/S026144480200188X

Sun, B., & Révész, A. (2021). The effects of task repetition on child EFL learners' oral performance. *Canadian Journal of Applied Linguistics*, *24*(2), 30–47. https://doi.org/10.37213/cjal.2021.31382

Suzuki, S. (2021). *A multidimensionality of second language oral fluency: The interface between cognitive, utterance, and perceived fluency*. [Unpublished doctoral dissertation]. Lancaster University.

Suzuki, S., & Kormos, J. (2023). The multidimensionality of second language oral fluency: Interfacing cognitive fluency and utterance fluency. *Studies in Second Language Acquisition*, *45*(1), 38–64. https://doi.org/10.1017/S0272263121000899

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

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

Suzuki, Y. (2023). *Practice and automatization in second language research: Perspectives from skill acquisition theory and cognitive psychology*. Routledge.

Suzuki, Y., Eguchi, M., & de Jong, N. (2022). Does the reuse of constructions promote fluency development in task repetition? A usage-based perspective. *TESOL Quarterly*, *56*(4), 1290–1319. https://doi.org/10.1002/tesq.3103

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

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

Thios, S. J., & D'Agostino, P. R. (1976). Effects of repetition as a function of study-phase retrieval. *Journal of Verbal Learning and Verbal Behavior*, *15*, 529–536. https://doi.org/10.1016/0022-5371(76)90047-5

Toppino, T. C., & Gerbier, E. (2014). About Practice: Repetition, spacing, and abstraction. In B. H. Ross (Ed.), *Psychology of learning and motivation* (Vol. *60*, pp. 113–189). Academic Press.

Westfall, J., Kenny, D. A., & Judd, C. M. (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *Journal of Experimental Psychology: General*, *143*(5), 2020–2045. https://doi.org/10.1037/xge0000014

Wiseheart, M., Küpper-Tetzel, C. E., Weston, T., Kim, A. S. N., Kapler, I. V., & Foot-Seymour, V. (2019). Enhancing the quality of student learning using distributed practice. In J. Dunlosky & K. Rawson (Eds.), *The Cambridge Handbook of Cognition and Education* (pp. 550–584). Cambridge University Press.

Yamagata, S., Nakata, T., & Rogers, J. (2022). Effects of distributed practice on the acquisition of verb-noun collocations. *Studies in Second Language Acquisition*, 1–27. https://doi.org/10.1017/S0272263122000225

Cite this article: Kakitani, J., & Kormos, J. (2024). The effects of distributed practice on second language fluency development. *Studies in Second Language Acquisition,* 1–25. https://doi.org/10.1017/S0272263124000251

https://doi.org/10.1017/S0272263124000251 Published online by Cambridge University Press