

Article

The Effects of Task Repetition Schedules on L2 Fluency Enhancement

Meng Zhang ¹, Na Yi ² and Dandan Zhou ^{1,*}¹ School of Foreign studies, Nanjing University, Nanjing 210023, China; zhangmeng2019@smail.nju.edu.cn² Department of Basic Courses, Engineering University of PAP, Xi'an 710086, China; yinanju327@163.com

* Correspondence: ddzhounj@nju.edu.cn

Abstract: This article explores the effects of different task repetition schedules on English learners' oral fluency in terms of speed, breakdown, and repair indices using a PRAAT script to enhance objectivity and consistency in the assessment of learner performances. A total of 90 freshmen participated in the experiment and were divided into three groups: the blocked repetition group, the interleaved repetition group, and a control group. This study adopted a pre- and post-test design. After the learners repetitively practiced the problem-solving tasks for three weeks, their improvement was measured by a new task of the same type. The analyses of speed, breakdown, and repair in learners' oral performance reported that the experimental groups outperformed the control group in fluency measurement. Specifically, the interleaved repetition group was advantageous compared with the blocked repetition group, except for the silent pause numbers. The findings not only support the interleaving effects and enrich the line of task repetition research but also have pedagogical implications in that introducing interleaved practice in real classrooms is beneficial to L2 speaking enhancement.

Keywords: task repetition schedules; blocked repetition; interleaved repetition; L2 oral fluency



Citation: Zhang, Meng, Na Yi, and Dandan Zhou. 2023. The Effects of Task Repetition Schedules on L2 Fluency Enhancement. *Languages* 8: 252. <https://doi.org/10.3390/languages8040252>

Academic Editor: Jon Clenton

Received: 28 June 2023

Revised: 12 October 2023

Accepted: 16 October 2023

Published: 25 October 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Learners' oral fluency, which is assumed to be one of the fundamental language abilities, serves as the crux of effective communication. Oral fluency has also been considered a critical aspect of language proficiency. China's Standards of English Language Ability (CSE) include fluency in language feature scales of oral expression, focusing on evaluation of speed, intonation, pauses, and so on. Being a fluent L2 speaker is also one of the ultimate goals that language learners dream of achieving. Nevertheless, oral fluency, in relation to processing ease and automaticity, is no mean feat. Experts in applied linguistics, language instructors, and practitioners spare no effort to search for how a language user can produce smooth, effortless, and automatic utterances. Moreover, the assessment of oral fluency is often labor-intensive and requires a high level of rater experience, which greatly limits the efficiency of related research. In today's technologically advanced world, technology-based fluency assessment can help measure students' performances and help teachers improve the efficiency, objectivity, and consistency of classroom evaluation.

1.1. Oral Fluency and Task Repetition Schedule

Oral fluency has been labeled as "procedural automaticity" in the psycholinguistic paradigm (Fillmore 1979). Oral fluency is considered as a language skill, the development of which is understood as increased automaticity. This skill-based view is supported by multi-disciplinary research, such as information processing and skill acquisition (Anderson 2009). Automatizing linguistic knowledge and skills inevitably links to practices of sufficient quantity and high quality (Suzuki 2022). Among various sorts of classroom activities to enhance oral fluency, task repetition has sparked widespread interest, which refers to "a task that has already been performed" (Bui et al. 2018, p. 227).

The importance of task repetition is highlighted by Segalowitz in that repetition functions as “a major route” to automatization (Segalowitz 2010), which is the prerequisite of utterance fluency and proficient L2 use. Research has revealed that task repetition assists in improving learners’ speaking fluency (e.g., Ahmadian 2011; Bygate 2001; de Jong and Perfetti 2011; Lambert et al. 2017; Thai and Boers 2016), which is perhaps due to the learners’ shift of attention from content conceptualization to formulation (Bygate 2001).

A recent line of task repetition inquiry focuses on task repetition schedules, among which the topic of blocked practice and interleaved practice has evoked substantial research interest (Suzuki 2021b). According to Suzuki (2021a), in a blocked condition, each of the practice tasks is repeated in order before moving on to the next task (e.g., AAABBBCCCC). While in an interleaved condition, tasks of various types are interspersed (e.g., ABCABCABC). It has been found that task repetition with different schedules may lead to different learning outcomes, for example, in grammar learning and speaking performance (Nakata and Suzuki 2019; Suzuki 2021b).

1.2. Theoretical Accounts of Task Repetition Schedules

The concept of interleaving and blocking is documented in several pedagogical frameworks, among which influential ones include the theoretical framework for systematic and deliberate L2 practice (Suzuki et al. 2019) and transfer-appropriate processing theory (Lightbown 2007). In order to identify optimal practice tailored for learners, Suzuki et al. applied the “desirable difficulty” concept to L2 practice and provided a theoretical framework that claims that instructors should take three key elements into account, including practice condition, linguistic difficulty, and individual features (Suzuki et al. 2019). In explaining context-related difficulty, spaced and interleaved practices are mentioned as effective distributed practices in the classroom. The concept of interleaving is also highlighted in the transfer appropriate theory (Lightbown 2007). Knowledge/skill transfer refers to “applying knowledge and skills that are acquired in one context to another” (Suzuki 2022, p. 326). The transfer-appropriate processing model focuses on this issue, and advocates that whenever a practice condition is similar to an outcome task, learning is transferred more readily (Lightbown 2007, 2019). The transfer-appropriate processing theory underlines how learning could transfer to the retrieval process instead of how learning takes place in the study phase.

The theoretical roots of interleaving effects can be traced back to cognitive science as well as educational psychology. Researchers proposed several theoretical bases for the effect of distributed practice, from the perspectives of discrimination, attention, and information encoding. The most popular of them is the discriminative contrast hypothesis, which asserts that interleaved presentations are more likely to enhance contrasts between items, making the differences more prominent (Kang and Pashler 2012). Two other explanations are proposed from the perspective of attention. One is the sequential attention theory, which argues that interleaving exercises can direct learners’ attention to different attributes of subsequent objects, thus increasing discrimination (Carvalho and Goldstone 2017). From the opposite angle, the attention attenuation hypothesis proposes that learners’ attention diminishes when they see repeated content, for example, the blocked presentation and practice (Zimmerman 1975). From the perspective of information encoding, the encoding-variability explanation favors greater variability in information presentation and believes it is conducive to information encoding (Gerbier and Toppino 2015).

1.3. Related Work on Task Repetition Schedules

To date, most of the results of related studies in language learning available support the interleaving effects on grammar, morphology, and syntax acquisition (Nakata and Suzuki 2019; Pan et al. 2019; Suzuki et al. 2022b). However, the effects of interleaved or blocked repetition on the development of skills, i.e., fluency training, are quite complex. In a recent study, English learners performed three oral storytelling tasks employing six-frame cartoons for three days in order to investigate the impact of task-repetition practice on fluency

development (Suzuki 2021b). They used either a blocked or an interleaved task repetition schedule to practice. The blocking group exhibited stronger fluency growth in some aspects (faster articulation rate and shorter pause duration) than the interleaving group, according to the results of a post-test using new six-frame cartoons. This study also revealed an intriguing “trade-off” relationship between pause frequency and pause duration, where learners unexpectedly created more pauses despite producing shorter pauses.

Suzuki (2021b) showcases contrasting findings with previous research which favors interleaving effects on L2 learning. The possible sources of the incongruence, according to the author, might be the nature of the practice, the difficulty or complexity of the employed task type, and the setting of multiple practice sessions. Interleaved practice, which has received recognition in psychology research, is seemingly less effective in the area of speaking practice.

A related study was conducted by de Jong and Perfetti, which investigated the impact of repeated task contents (or no-repeated contents) on L2 fluency development and proceduralization (de Jong and Perfetti 2011). Although de Jong and Perfetti did not refer their study design to the specific term, the topic-repetition group in their study corresponded to the blocked practice. Students learning English as a second language at a university in the United States participated in nine 4/3/2-min monologues, with the time given for each task decreasing over the course of two weeks. The no-topic-repetition group performed the speaking task on nine distinct themes (training session 1: topics 123; training session 2: topics 456; training session 3: topics 789). The topic-repetition group performed the speech task three times on the same topic, which met the requirement of blocked practice (training session 1: topic 111; training session 2: topic 444; training session 3: topic 777). To see if the fluency benefits from repeated practice might be transferred to speaking performance on a different topic, one-week and four-week delayed post-tests were given. Only the repetition group (blocked practice) showed notable fluency growth linked to proceduralization (e.g., mean length of run, pause length, and phonation/time ratio) when the delayed post-tests were examined.

These two studies paved the way for exploring blocking/interleaving effects on L2 speaking fluency. They indicate the possibility of promoting proceduralization by manipulating practice distribution and sequence. Suzuki used the term “practice variability” to capture the possible factor that impacts underlying L2 speaking mechanism and proceduralization (Suzuki 2022). The two pioneering studies aforementioned seem to indicate that a less variable condition is advantageous to fluency transfer. However, the benefits of a task repetition schedule (blocking or interleaving) depend on the learning materials, task types, and skills (Brunmair and Richter 2019). Different task types being observed (for example, those with higher cognitive demands) have received scant attention in this line of research. Moreover, there are still questions untouched. Both studies were conducted in the laboratory. The investigation should go beyond the lab-based setting because it is not clear whether task repetition schedules take effect in real classroom settings. Up to now, far too little attention has been paid to the different effects of blocked and interleaved repetition on complex skill development, for example, L2 fluency.

Employing problem-solving tasks, the following research questions guided the current study:

- (1) To what extent do task repetition schedules impact L2 fluency (speed, breakdown, and repair) compared to the control group?
- (2) What is the relative effectiveness of task repetition schedules (blocking and interleaving) on participants’ oral fluency in terms of speed, breakdown, and repair?

2. Materials and Methods

2.1. Participants

A total of 90 non-English majors in their freshman year took part in the current study at a university in northwest China. They were from six intact classes. The six classes were randomly assigned to either experimental groups (two classes in the blocked repetition

group and two classes in the interleaved repetition group) or the control group (two classes receiving no training). The participants were instructed by the same English teacher. They took the Oxford Quick Placement Test to assess their English proficiency levels. One-way ANOVA reported that there were no significant differences between the participants among the three groups ($p = 0.704$). They signed a consent form prepared following the ethical guidelines. The experiment was carried out in their regular listening and speaking classes. During the experiment, two of the participants in the control group failed to complete all tasks and were thus excluded from the data analysis.

2.2. Instruments

2.2.1. Oral Task

The present study employed the modified tasks that were detailed for Lambert et al., using a structured problem-solving task (“The Abby Task” by Lambert et al. 2020). “The Abby Task” presented the scenario that the character Abby (played by the students themselves) receives several “pour out letters” about the troubles the letter-writers encountered in their campus life. The students were required to reply to the letter-writers orally, following a detailed problem-solving structure. There are three major advantages of employing the Abby tasks. First, it is in accordance with students’ campus life, which invites their engagement. Second, it provides a battery of parallel oral tasks with similar task complexity, which are selected to ensure compatibility with each other. It is suitable for experiments with a repetition schedule because it can largely control the inner differences of each task, for example, familiarity and complexity, which may obscure the main effects of a repetition schedule. Third, the tasks have a clear structure for participants to follow, which can address students’ difficulty organizing content and is useful for training their thinking and making their speech more logical.

The topics for the pre- and post-tests included “keeping balance between part-time job and schoolwork” and “getting out of the haze of lost love”. The topics for practices were “reminding friends to pay back the money” (Task A), “convincing my parents to allow me to study abroad” (Task B), and “getting up early to attend classes at a distant campus” (Task C). Students produced monologues individually, which were recorded by computers. Prompts were displayed on the screen with a different headshot in each iteration to create a novel “interlocutor”, which was intended to relieve boredom. The participants were asked to complete each task within 2.5 min and follow the problem-solving procedure: explaining the problem, comparing possible solutions, recommending one of these, and giving the reason for the recommendation. This procedure helps to ensure the parallel nature of the tasks. More information related to the prompts was provided in Appendix A. Participants had 2 min to prepare for each task, and they were allowed to take notes on phrases, but not complete sentences.

2.2.2. Questionnaire

The current study modified Cho’s survey of examining learner experience (Cho 2018), including the perception of task difficulty, skill, and flow (interest, attention, and control), with background information added. This questionnaire was chosen for the following reasons: First, the focused items are suitable for the present study. The most typical student assessment of a task is its perceived difficulty (Cho 2018). Moreover, task difficulty has traditionally been viewed as an add-on construct that confirms the methodological operationalization of task features (Tavakoli 2009). Additionally, the appropriate challenge level of a task, an important component of flow, ties into the present discussion of learner experience. It also echoes The Desirable Difficulty Framework (Suzuki et al. 2019), which searches for the optimal challenge for language practice and invites discussion about the synergy among practice conditions and learner-related difficulty factors. Second, it contains a moderate number of items, which allows repeated measurement. Cho had the participants complete the questionnaire four times (one for each task, Cho 2018). The present study distributed the questionnaire at the end of each training session (three times in total).

2.3. Procedures

This study was conducted in dedicated listening and speaking classrooms equipped with integrated digital language learning system DH500E-Ningdayifang-NDE2. Additionally, high-quality audio recording capabilities were ensured through the use of Panasonic's hardware equipment, WE-LL310A. These systems allowed for precise control over each computer, as well as the corresponding headphones and microphone setups, ensuring high-quality audio recording. Participants were seated at intervals in these classrooms, with each learner surrounded by an empty space to minimize the influence of neighboring participants on the audio recordings. All participants used their dedicated headphones and microphone setups corresponding to their seats. The mitigation of background noise influence is further controlled through parameter adjustments and manual verification. Detailed information is reported in the data analysis section.

The current study adopted a pretest-treatment-post-test design. The procedure is presented as follows (Table 1).

Table 1. Experiment procedures.

Group	Pretest (Week 1)	Treatment (Week 2, 3, 4)	Post-test (Week 5)
Blocked group	Pre-test	AAA-BBB-CCC	Post-test
Interleaved group		Q&D-Q&D-Q&D	
Control group		ABC-ABC-ABC Q&D-Q&D-Q&D /	

Q: questionnaire; D: diary; The topics for pre- and post-tests were counterbalanced.

Topics for pre- and post-tests were employed in counterbalanced orders to eliminate any order effects. All the participants were assigned to either a blocked-practice condition ($n = 30$) or an interleaved-practice condition ($n = 30$). The learners in the control group ($n = 30$) took the pretest and post-test only. The experiment was conducted in the first session of their listening and speaking course each week. Each practice session (three oral tasks) was immediately followed by a survey and a diary recording the students' experiences with the task.

2.4. Measures

2.4.1. Measures of Fluency

Oral fluency is considered as overall competence in a broad sense (Tavakoli et al. 2016). While in a narrow sense, fast speech, as well as the relative absence of unnecessary hesitations, halting, repetition, and self-repairs, are features that most people associate with fluency (Segalowitz 2010). The current research took the latter view, focusing on speed, breakdown, and repair fluency. Based on Skehan's taxonomy of utterance fluency (speed fluency, breakdown fluency, and repair fluency, Skehan 2003), each dimension of fluency was analyzed as in Table 2 with reference to the prior studies on fluency (Lambert et al. 2020; Suzuki 2021b; Tran and Saito 2021).

It should be acknowledged first that there have been no perfect or totally systematic measurements so far (Tavakoli and Wright 2020). In the current study, specific measures were selected with due consideration of the following advantages: First, the current study included both pure measures and composite ones for different aspects of L2 fluency in order to take a more systematic approach to measurement (Tavakoli and Wright 2020), and the multidimensional nature of fluency research requires more refined and deeper evaluation (Suzuki and Kormos 2023). While composite measures such as speech rate and mean length of run are the most representative indices that could capture fluency generally (Hunter 2017), the pure measures are also contributive since it may provide evidence of changes in different stages in oral production (Suzuki and Kormos 2023). For instance, frequency of silent pauses is associated with conceptualization and formulation stage, while

articulation rate is assumed to be related to articulation stage in the L2 oral production model (Tavakoli and Wright 2020). Second, the incorporation of ratios (phonation/time ratio and pause/time ratio) is a response to Tavakoli and Wright’s call for measuring fluency on the basis of ratios so as to increase comparability across studies (Tavakoli and Wright 2020). Third, the interpretation of mean length of run, phonation/time ratio, and mean length of pause are believed to indicate the tendency toward proceduralization (de Jong and Perfetti 2011), which plays a significant role in fluency development.

Table 2. Measures of fluency.

Fluency Aspects		Operationalization
Speed fluency	Speech rate (SR)	The number of syllables per minute of speech (including pause)
	Articulation rate (AR)	The number of syllables per minute of speaking time (excluding pauses)
	Phonation/time ratio (PTR)	Utterance duration divided by the total duration
	Mean length of run (MLR)	Total numbers of syllables divided by total number of runs
Breakdown fluency	Frequency of silent pauses (FSP)	Number of silent pauses divided by the total duration, multiplied by 60
	Pause time ratio (PATR)	Total silent pause duration divided by total duration
	Mean length of pauses (MLP)	Total silent pause duration divided by the numbers of pauses
Repair fluency	Repetition frequency (RF)	The number of repetitions per minute
	Self-repair frequency (SRF)	The number of self-repairs per minute

It should also be noted that the classification of these measurements is not exclusive but follows the majority of empirical studies. For example, according to the interpretation of Tavakoli and Wright (2020), the articulation rate is a pure measurement of speed fluency, while the mean length of run is a composite measurement because it reflects both the roles of silence frequency and syllables produced. The measurement of speech rate even incorporates the three fluency aspects. Repair, pause phenomena, and the syllables produced all contribute to the final value. Moreover, some research holds that repetition in repair fluency may indicate “stalling”, which falls into the scope of breakdown fluency (Dörnyei and Kormos 1998). Since inconsistent views still exist on this issue, we chose to follow most of the studies in which repetitions are regarded as repair behaviors.

With the help of a PRAAT script (de Jong and Wempe 2009), pauses of at least 250 ms duration were identified using the free sound-analysis software PRAAT (Version 6.0.14, Boersma and Weenink 2016), and indices such as speech rate, articulation rate, phonation/time ratio, and pauses-related measurements were calculated automatically by the software and the widely acknowledged script (e.g., Suzuki 2021a; Suzuki et al. 2022a). Automatic scoring by the script has been proven to highly correlate with manual measures (>0.8) (de Jong and Wempe 2009). Using automatic identification of syllables and pauses has several merits. To begin with, compared to holistic human rating, automated analysis is more objective and ensures consistency. According to Tavakoli and Wright, the past two decades have witnessed a methodological shift, and digital software can now report more precise and efficient results, especially in the dimension of temporal fluency (Tavakoli and Wright 2020). Moreover, in the current study, three strategies were employed to increase the validity of fluency assessments. First, the identification of the pause threshold followed de Jong and Bosker’s recommendation of 250 ms. It is believed that native speakers have a negative perception of fluency with longer pauses (longer than 250 ms), and this threshold can best distinguish pauses (de Jong and Bosker 2013). The commonly used threshold of 250 ms is also selected in accordance with previous empirical research (e.g., Bao 2023; Saito et al. 2022; Suzuki and Kormos 2023). Second, the silence threshold was set to −20 dB, which is supposed to avoid noise interference and fit into the classroom setting (Saito et al. 2022). Third, manual re-checking was carried out after

automatic identification of pauses and syllables. A total of 3.34% of the annotations were adjusted, with an inter-rater reliability of 0.93 between the two coders.

The first and second authors transcribed all the speaking samples (88×2) and counted the number of repetitions and self-repairs independently. They discussed and reached consensus when disparities appeared in the identification of repetitions and self-repairs.

Example of repair coding (Student B-21):

First, I know you really want to change the change the (repetition *1) condition situation (repair *1), change this situation (repair *1). So you can, so you can (repetition *1) developments develop (repair *1) a new relationship with others.

2.4.2. Measures of Learner Perception and Experience

The present study used Cho's questionnaire, tapping into three aspects of learner experience: perceived difficulty, perceived skill, and flow experience, with 15 items in total (Cho 2018). On a 6-point Likert scale, participants were asked to score the statements' appropriateness, with 1 indicating "not at all appropriate" and 6 indicating "extremely appropriate". In terms of perceived difficulty, learners evaluated levels of "difficult" and "easy" in two separate items; thus, the score of "easiness" was reversely coded. In terms of perceived skill, 3 items centering "ability" and "competence" were given marks. As to interest, 3 items concerned about the intrinsic enjoyment and the willingness to participate in the practice again. As to attention, 3 items addressed the extent to which learners were engaged in practices without being distracted by other things. As to control, 4 items investigated whether the goal of the practice was clear and whether learners felt in control. Within each construct, the scores were averaged.

2.5. Data Analysis

To investigate between-group differences in speed, breakdown, and repair behaviors, univariate ANCOVAs were conducted, following Suzuki's design, when comparing between different conditions (Suzuki 2021b). The reason to choose ANCOVAs is that they could control potential differences for the first time. The focus lies on changes from the pre-test to the post-test between different groups. Moreover, the design, which incorporated a control group, could control potential sources of confounding variables like the opportunities to practice oral English in their language courses because the participants in the control group took the same English course with the same instructor as the two experimental groups. Three levels of groups (blocked, interleaved, and the control group) were the between-subject variables. The post-test score was the dependent variable, and the corresponding pre-test score was the covariate in each fluency measure.

As to the data from questionnaires, two-way repeated-measures multivariate analysis of variance (MANOVA) were conducted. The within-subject variables were three levels of week and two levels of condition as the between-subject variables, and the dependent variables (DVs) were perceived task difficulty, skill, and flow. Since learner experience was not the focus of the current study, the whole bunch of results would not be included in the result section. The related findings were used only when explaining the possible reasons underlying the fluency change in the discussion part. Learners' diaries were also consulted to provide qualitative evidence.

3. Results

By assessing Q-Q plots and skewedness and kurtosis, dependent variables were approximately normally distributed except for MLR, RF, and SRF. Thus, a reciprocal transformation was conducted to correct the distribution of MLR. A square root transformation was conducted on repair measures. The current study followed the benchmark of effect sizes specific to blocking or interleaving effects ($\eta_p^2 > 0.06$; Cohen's $d > 0.40$), with references to the previous research (Suzuki 2021b).

3.1. Speed Fluency

According to Table 3, there were significant differences among the three groups in speed fluency measures (SR, $p = 0.000$, $\eta_p^2 = 0.218$; AR, $p = 0.000$, $\eta_p^2 = 0.167$; PTR, $p = 0.000$, $\eta_p^2 = 0.188$; MLR, $p = 0.008$). Pairwise comparisons (Table 4) detected the significant differences in SR between the blocked repetition group and the control group ($p = 0.034$) and the interleaved repetition group and the control group ($p = 0.000$). Both the experimental groups outperformed in SR than the control group. The difference between the blocked and interleaved groups was meaningful (Cohens' $d = 0.49 > 0.40$), with the interleaved group performing better. As to AR, significant differences were found between the blocked and the interleaved repetition groups ($p = 0.029$), and between the interleaved repetition group and the control group ($p = 0.000$), with the interleaved group producing a higher AR compared with the other groups. In terms of PTR, both the blocked repetition group ($p = 0.007$) and the interleaved group ($p = 0.000$) did better than the control group. With regard to MLR, significant differences were noted among groups ($p = 0.008$), with both experimental groups outperforming the control group ($p = 0.023$, 0.019 respectively).

Table 3. ANCOVA results of speed fluency adjusted for pre-test.

Measures	Group	Raw Mean	Adjusted Mean	Std. Error	F	<i>p</i>	η_p^2
SR	B	126.56	123.03	5.79	11.73	0.000 *	0.218
	I	146.22	142.23	5.80			
	C	93.00	101.06	6.10			
AR	B	220.76	216.58	4.67	8.45	0.000 *	0.167
	I	231.00	234.19	4.65			
	C	206.19	207.26	4.80			
PTR	B	0.57	0.57	0.02	9.70	0.000 *	0.188
	I	0.63	0.61	0.02			
	C	0.45	0.47	0.02			
MLR ^	B	0.19	0.19	0.01	5.05	0.008 *	0.107
	I	0.18	0.18	0.01			
	C	0.23	0.23	0.01			

$df = 2$; B: blocked repetition group, $n = 30$; I: interleaved repetition group, $n = 30$; C: control group, $n = 28$; * $p < 0.05$; ^ reciprocal transformation was conducted on MLR, therefore a smaller value indicates a higher mean length of run.

Table 4. Pairwise comparison for speed fluency.

Measures		Mean Differences.	Std. Error	<i>p</i>	95% IC	Cohen's <i>d</i>
SR	B vs. I	−19.21 +	8.15	0.062	[−39.13, 0.71]	−0.61 +
	B vs. C	21.97 *	8.49	0.034	[1.23, 42.70]	0.70 +
	I vs. C	41.18 *	8.50	0.000	[20.41, 61.94]	1.30
AR	B vs. I	−17.61 *	6.64	0.029	[−33.82, −1.40]	−0.69 +
	B vs. C	9.33	6.71	0.504	[−7.06, 25.71]	0.37
	I vs. C	26.93 *	6.67	0.000	[10.64, 43.23]	1.06 +
PTR	B vs. I	−0.04	0.03	0.619	[−0.11, 0.04]	−0.33
	B vs. C	0.10 *	0.03	0.007	[0.02, 0.17]	0.83 +
	I vs. C	0.14 *	0.03	0.000	[0.06, 0.21]	1.15 +
MLR ^	B vs. I	0.002	0.02	1.00	[−0.04, 0.04]	0.03
	B vs. C	−0.05 *	0.02	0.023	[−0.09, −0.01]	−0.79 +
	I vs. C	−0.05 *	0.02	0.019	[−0.09, −0.01]	−0.79 +

Pairwise comparison: Bonferroni correction; * $p < 0.05$, + meaningful effect size. ^ reciprocal transformation was conducted on MLR, therefore smaller values indicate longer MLR.

3.2. Breakdown Fluency

As shown in Table 5, ANCOVAs reported significant differences among the three groups in PATR ($p = 0.000$, $\eta_p^2 = 0.188$) and MLP ($p = 0.000$, $\eta_p^2 = 0.183$). Although the

differences in FSP did not reach the significance level between groups FSP ($p = 0.093$, $\eta_p^2 = 0.055$), the blocked group produced less FSP relative to the control group with a meaningful effect size (-0.54), as reported in Table 6. However, the interleaved group produced more FSP compared with the control group, with a meaningful effect size (0.45). Pairwise comparison detected that both the blocked ($p = 0.007$) and the interleaved groups ($p = 0.000$) produced significantly smaller PATR than the control group. In terms of MLP, both the blocked ($p = 0.027$) and the interleaved repetition group ($p = 0.000$) spoke with shorter MLP compared with the control group. Although the difference between the two experimental groups did not reach the significance level, it was meaningful with an effect size of 0.48 . The interleaved group produced a shorter MLP than the blocked group.

Table 5. ANCOVA results of breakdown fluency adjusted for pre-test.

Measures	Group	Raw Mean	Adjusted Mean	Std. Error	F	<i>p</i>	η_p^2
FSP	B	21.53	21.22	0.93	2.45	0.093	0.055
	I	23.82	23.93	0.92			
	C	21.46	21.68	0.96			
PATR	B	0.43	0.43	0.02	9.70	0.000 *	0.188
	I	0.37	0.39	0.02			
	C	0.55	0.53	0.02			
MLP	B	1.30	1.34	0.11	9.42	0.000 *	0.183
	I	0.97	1.06	0.11			
	C	1.91	1.78	0.12			

df = 2; B: blocked repetition group, *n* = 30; I: interleaved repetition group, *n* = 30; C: control group, *n* = 28; * $p < 0.05$.

Table 6. Pairwise comparison for breakdown fluency.

Measures		Mean Differences	Std. Error	<i>p</i>	95% IC	Cohen's <i>d</i>
FSP	B vs. I	-2.72^+	1.32	0.126	$[-5.93, 0.50]$	-0.54^+
	B vs. C	-0.47	1.34	1.000	$[-3.75, 2.82]$	-0.09
	I vs. C	2.25^+	1.33	0.282	$[-1.00, 5.50]$	0.45^+
PATR	B vs. I	0.04	0.03	0.619	$[-0.04, 0.11]$	0.33
	B vs. C	-0.098^*	0.03	0.007	$[-0.17, -0.02]$	-0.83^+
	I vs. C	-0.136^*	0.03	0.000	$[-0.21, -0.06]$	-1.15^+
MLP	B vs. I	0.28^+	0.16	0.241	$[-0.11, 0.67]$	0.46^+
	B vs. C	-0.440^*	0.16	0.027	$[-0.84, -0.04]$	-0.72^+
	I vs. C	-0.720^*	0.17	0.000	$[-1.13, -0.31]$	-1.18^+

Pairwise comparison: Bonferroni correction; * $p < 0.05$, + meaningful effect size.

3.3. Repair Fluency

According to Table 7, three groups showed significant differences in both indices (RF, $p = 0.011$; SRF, $p < 0.001$). Pairwise comparisons (Table 8) showed that the blocked group presented significantly fewer repetition behaviors compared with the control group ($p = 0.006$). The interleaved group also significantly outperformed the control group in RF ($p = 0.001$). As to SRF, the interleaved group did better than the blocked group ($p = 0.010$) and the control group ($p < 0.001$) significantly.

Table 7. ANCOVA results of repair fluency adjusted for pre-test.

Fluency Measure	Group	Raw Mean	Adjusted Mean	Std. Error	df	F	p	η_p^2
RF	B	0.88	0.84	0.08	1.55	7.89	0.001 *	0.158
	I	0.74	0.78	0.08				
	C	1.21	1.21	0.08				
SRF	B	0.82	0.82	0.08	2	10.71	<0.001 *	0.203
	I	0.48	0.49	0.08				
	C	1.03	1.01	0.08				

df = 2; B: blocked repetition group, n = 30; I: interleaved repetition group, n = 30; C: control group, n = 28; * $p < 0.05$; A square root transformation was conducted.

Table 8. Pairwise comparison for repair fluency.

Measures	Mean Differences	Std. Error	p	95% IC	Cohen's d
RF	B vs. I	0.06	1.00	[−0.23, 0.34]	0.14
	B vs. C	−0.37 *	0.006	[−0.66, −0.09]	−0.83 +
	I vs. C	−0.43 *	0.001	[−0.72, −0.14]	−0.97 +
SRF	B vs. I	0.33 *	0.010	[0.06, 0.60]	0.78 +
	B vs. C	−0.19	0.307	[−0.46, 0.09]	−0.45 +
	I vs. C	−0.52 *	<0.001	[−0.79, −0.24]	−1.22 +

Pairwise comparison: Bonferroni correction; * $p < 0.05$; + meaningful effect size; A square root transformation was conducted.

To summarize, the experimental groups overmatched the control group in speed fluency (SR, PTR, MLR), breakdown fluency (PATR, MLP), and repair fluency (RF). In SR, AR, MLP, and SRF, the interleaved group surpassed the other two groups, thus becoming the optimal practice condition in these aspects. However, in FSP, it was a suboptimal choice. The two experimental groups were comparable in PTR, MLR, PATR, and RF.

4. Discussion

Findings yielded by the statistical analyses indicated that both the blocked and the interleaved repetition groups had better performances in L2 fluency as measured by the post-test. Between the two, the interleaved repetition group presented an advantageous output in terms of fluency. In what follows, these results are discussed in relation to theoretical accounts, and evidence from participants' performances, questionnaires, and diaries pertaining to their practice experience and perception.

4.1. Merits of the Experimental Groups

Findings revealed that both experimental groups outperformed the control group in speed, breakdown, and repair fluency. This observation could possibly be accounted for transfer-appropriate processing theory, emotional engagement in practice, and proceduralization to some extent.

DeKeyser emphasizes the importance of transferability of skills in that automatization not only contains the “speed-up of the same basic mechanisms” but also the “speed-up of a broader task” (DeKeyser 2015, p. 96). Yet, to achieve transferability, there are conditions to meet. Findings yielded by prior research indicate that blocking (content repetition within practice sessions) has the potential to facilitate fluency transfer (de Jong and Perfetti 2011; Suzuki 2021b). However, in the current study, both the blocked and interleaved repetition groups behaved more fluently in the new task (post-test) compared with the control group. These different sets of results may stem from the types of tasks used in each study. De Jong and Perfetti deployed monologue on various topics ranging from “opinions toward pets” to “a favorite artist” (de Jong and Perfetti 2011). The topics are less similar intrinsically. This dissimilarity hinders interleaving from taking effect since high similarity between tasks is documented to enhance interleaving (Brunmair and Richter 2019). The same is true in Suzuki's study (Suzuki 2021b). Although the picture-story tasks were also structured

following the lead of plots, the tasks being practiced (bicycle, tiger, and race) differed from each other in conceptualization since the contents were relatively fixed and restricted. Because the Abby tasks used in the current study during the practice and in the pre- and post-tests resembled each other in their tight problem-solving structure, this allowed interleaved practices to take effect and provided easier access to transferability.

The similarity between tasks being practiced and tested provides favorable conditions for transfer. Learning is transferred more readily whenever a practice condition is similar to an outcome task (Lightbown 2007). This theory has two implications. First, the practice session and the test session in the study itself should be similar. Second, the tasks used in the practice should be similar to students' practice in daily life. Both conditions are crucial to the successful transfer. In contrast to the two pioneer studies (de Jong and Perfetti 2011; Suzuki 2021b), Hunter (2017) failed to detect L2 fluency transfer because it employed a "poster carousel" task during practice sessions while using topic-development tasks in the pre- and post-tests. She attributed the results to the violation of the transfer-appropriate theory. Practice conditions in the current study also fit into appropriate transfer in that the chosen topics (background information about the problems to be solved) are close to students' campus life. The problems are what contemporary college students encounter and search for solutions for, such as the pain of getting up early for class across campus, the struggle of reminding friends to return money, and torturous lovelorn time. In keeping with the goal of promoting communicative skills in actual use, practices involving tasks close to reality should be considered appropriate in light of the similarity between controlled context (in the classroom) and actual communication.

Drawing on the desirable difficulty framework, the current findings may also benefit from the balance between condition and learner-related difficulty. Suzuki regarded the training content in his study as too challenging for the participants, thus making the interleaved practice more difficult, overwhelming them, and adversely affecting proceduralization development (Suzuki 2021b). In the current study, students from both experimental groups completed questionnaires three times per week during the practice sessions. The within-group variable was time, the between-group variable was group, and the dependent variables were difficulty, skill, interest, attention, and control. Evidence from the questionnaire showed that the students felt tasks were significantly less difficult ($F(2, 116) = 19.44$, $p < 0.001$, with means being 3.84, 3.36, 3.28 from week 1 to week 3) and their skills were improving during the practices ($F(2, 116) = 16.72$, $p < 0.001$, with means being 3.30, 3.69, 3.88 from week 1 to week 3) in the two experimental groups when examining the main effects of time. Moreover, emotional engagement plays a vital role in L2 skill transfer (Lightbown 2019). In order to detect whether learners positively engaged in the practices, the data from the questionnaire were resorted to. According to the participants' self-assessment, despite the stability of interest ($F(2, 116) = 0.12$, $p = 0.885$), other variables like attention ($F(2, 116) = 3.96$, $p = 0.022$, with means being 3.97, 4.08, 4.27 from week 1 to week 3) and control ($F(2, 116) = 8.51$, $p < 0.001$, with means being 3.73, 4.08, 4.16 from week 1 to week 3) were getting better as the practices progressed in the two experimental groups when examining the main effects of time. It is speculated that positive engagement might be one of the factors that enhances the practice effect and transferability.

Drawing on previous research, systematic practice covering three weeks might change the linguistic encoding stages (de Jong and Perfetti 2011), even at the structure building at the clause and sentence level (Kormos 2006). Although problems in the training sessions were different from those in the post-tests, the common component (e.g., similar task structure) allowed learners to proceduralize certain expressions, for example, the linguistic structure of providing and evaluating solutions, and even pragmatically communication routines.

There is also the possibility that fluency improvement in the post-test relative to the pre-test profits from the proceduralization of abstract constructions. Prior research speculates that reusing trigrams of part of speech (POS) promotes oral fluency development through the proceduralization of abstract constructions (Suzuki et al. 2022a; de Jong and

Tillman 2018). In task repetition, certain constructions were reused. Thus, in subsequent performances, the retrieval of these constructions was accelerated. After a period of practice, these constructions were able to be entrenched in memory. Repeated use of certain constructions can reinforce the retrieving traces and boost retrieval efficiency, leading to linguistic proceduralization. Repeated practices invite similar propositional expressions and increasingly activate and entrench related abstract constructions, the process of which strengthens proceduralization to some extent.

4.2. Comparative Advantage of Interleaving

Interleaving in the current study proved to be more effective than blocking. The finding might be explained by the theory of sequential attention and encoding variability. The sequential attention theory posits that interleaved practices can direct learners' attention to different attributes of subsequent objects, thus strengthening discrimination and coping strategies (Carvalho and Goldstone 2017). Correspondingly, learners' attention diminishes when they see repeated content, for example, in blocked presentation and practice (the attention attenuation hypothesis, Zimmerman 1975). There is evidence from the questionnaire. During practice session 1, the mean number of participants' attention was 4.05 in the interleaved group, with a mean of 3.90 in the blocked group. In practice sessions 2 and 3, the interleaved group was measured as paying more attention to the tasks compared with its counterpart (week 2: 4.09 vs. 4.08; week 3: 4.42 vs. 4.12). Participants in the blocked repetition group also acknowledged attenuation of attention during practice, such as participants B-07 and B-26 in their diaries:

Example 1: *"I was nervous when time was running out, but sometimes I felt my mind was wandering". (Participant B-07)*

Example 2: *"Sometimes the lack of concentration may lead to mechanical repetition and knowledge did not enter the brain". (Participant B-26)*

In the blocked group, attention attenuation during practices may weaken the practice effects to some extent. On the one hand, the participants benefited from lexical or sentence priming during the same task repetition within each practice session. Their attentional resources were freeing up in the subsequent performances (at T2 and T3). Fulfilling the following tasks did not require as many attentional resources as previously. The fluency change seemed positive in the blocked repetitions during practices. On the other hand, the practice effects may be leveled off by attention attenuation. Since performance was different from learning, the ease with which the participants performed tasks during practices did not guarantee favorable results in testing measured by a new task. On the contrary, the participants in the interleaved repetition group who put in relatively more effort with high concentration would gain bonuses in testing, though sometimes their performances were not outstanding during practices.

Drawing on information encoding theory, the encoding–variability explanation favors greater variability in information presentation and believes it is conducive to information encoding (Gerbier and Toppino 2015). Variation in contextual cues causes encoding variability. A change in the context, for example, various situations and problems encountered in the tasks in the interleaved practice, may reduce the recall of information at present (interleaving may be inferior to blocking during practice); however, a greater range of contextual cues will be attached to it. As a result, a greater level of encoding variability facilitates maintaining access to information across different contexts, thus fostering learning transfer. On the contrary, the benefits of learning may be reduced when participants in the blocked repetition group exclusively focus on the same context and problem without accounting for contextual cues during the encoding process. The contrast between pause frequency and pause duration in the interleaved repetition group may also be influenced by encoding variability during practices. The participants were required to generate various

ideas to deal with the different problems they encountered. They may habitually pause to conceptualize. However, the length of pauses was reduced in the post-test, indicating the processing is speeding up.

In relation to encoding variability, closed and open skill accounts can contribute to the current discussion. As to the distinctions between skills practiced during blocked and interleaved repetition, they could be attributed to differences in processing flexibility (Segalowitz 2010). Related concepts are closed skills and open skills. A closed skill refers to a physical or cognitive act that is performed in a relatively stable environment and aims to reproduce more precisely as possible an ideal version of the act, while open skill denotes skills that take place in uncertain environments where the goal of performance is to affect the environment in some way (Allard and Starkes 1991). Segalowitz reminded researchers that closed skills and open skills are not exclusive from each other and that it is common to find skills containing both of them, and the importance of the distinction between them invites researchers' attention to how learners recruit diverse processes that drive performance for fluent skill execution (Segalowitz 2010).

Learners in the current study practiced more of the open skills under the interleaved repetition condition within each practice session because they encountered different situations (A, B, C) whereas learners under the blocked repetition condition dealt with the same problem within each practice session (e.g., A, A, A). Closed and open skills have different cognitive processing requirements for second language learners (Segalowitz 2010), so within the practice session, tasks were more stressful for the interleaved group to handle online, but the training was more intense for the open skills, perhaps enhancing the ability to handle new problems in new environments (fluency transferability). In the blocked repetition group, practices within each practice session drove more of the closed skill performance, involving a high degree of predictability in the situation. However, L2 use is more like open skills as opposed to closed skills, and transfer requires training to deal with unpredictability (Segalowitz 2010). Thus, in the final test, faced with a new problem, the participants in the interleaved group were relatively superior to the blocked group.

As has been mentioned in Section 4.1, the findings in the current study contradicted Suzuki's study, where interleaving failed to take effect (Suzuki 2021b). The possible reasons could be the different tasks employed and some differences in the research design, such as the intervals between practice sessions and the experiment settings. As to task types, Suzuki's study used the picture-story tasks while the current study employed problem-solving tasks, which may have different cognitive demands for the participants. For example, the problem-solving tasks might require more cognitive effort in conceptualization, such as providing viable solutions, but the learners might be more flexible in content according to their own linguistic resources. However, the content of learners' oral production was relatively restricted to the pictures. Thus, the variability of encoding might be more pronounced in the present study. Moreover, as discussed before, the similarity between practice and testing is considered important for interleaving practice to take effect (Brunmair and Richter 2019). The highly structured nature of the problem-solving tasks might also contribute to the current findings. As to different research designs, the current study set the interval between each session as one week, while that in Suzuki's study was one day. Some measures of fluency, i.e., repair fluency, were found to be more susceptible at a one-week interval in the previous study (Bui et al. 2019). It might be attributed to the adequate opportunity rendered by such a spacing for practice of encoding and retrieval. Another difference in research design is related to the experiment settings. Learners in their authentic classes might be less anxious, thus being more likely to embrace variability during practice. To sum up, the results in the present study added more empirical evidence to the field of interleaved practice in second language learning.

5. Conclusions

The results of the current study suggest that repetitive practice enables the development of oral fluency. Moreover, different task repetition schedules may influence fluency

development, with interleaving repetition resulting in more favorable behaviors in utterance speed, pause duration, and repair phenomena. In terms of less frequent pauses, blocking repetition was advantageous. The current study supports the idea that manipulating task repetition schedules could impact learner behaviors of fluent speaking and provides empirical evidence of interleaving effects on L2 speaking training. It also contributes to implementing frameworks of desirable difficulty and transfer appropriate processing in real classroom settings. Pedagogically, the current study provides some enlightenment for systematic repetition, which intends to foster fluency development in classroom settings. The influence of laboratory and classroom settings on students may vary. This study was conducted in authentic listening and speaking classrooms, where the completion of corresponding speaking tasks was integrated into the regular course curriculum. Students in the classroom setting may be more relaxed compared to laboratory-based research (Hunter 2017). Hence, the findings of the current study may be more inspirational compared to laboratory-based research if the goal is to provide more empirical evidence for pedagogical practices.

This study was completed within one month. Therefore, further longitudinal research is needed. Due to the difficulty of analyzing fluency performance, the number of participants was limited. Future studies may incorporate more participants with various educational backgrounds and investigate the moderating role of other learner factors in the effects of the repetition schedule, such as L2 proficiency, language aptitude, and L1 fluency style. Moreover, the measurements of L2 oral fluency have long been complex. Existing software and scripts are still not capable of probing some indices of fluency, such as the location of pauses, which is absent in the current study. As technology continues to innovate, the problem of automatically evaluating these metrics should be solved in the near future. Finally, to control experimental variables, this study chose monologues. However, it should be acknowledged that dialogues offer greater interactivity, align more closely with learners' needs in daily life, and are better at reducing boredom during practice. Future research should further investigate the use of dialogues such as 4-3-2 activity which incorporates a real interlocutor.

Author Contributions: Conceptualization, M.Z. and D.Z.; methodology, M.Z.; software, M.Z.; validation, M.Z., N.Y. and D.Z.; formal analysis, M.Z. and N.Y.; investigation, N.Y.; resources, M.Z.; data curation, M.Z.; writing—original draft preparation, M.Z.; writing—review and editing, N.Y. and D.Z.; supervision, D.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Ethics Committee of School of Foreign Studies, Northwest University of Political Science and Law (date of approval: 15 November 2021).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on reasonable request from the first author.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. The Abby Tasks (Adapted from Lambert et al. 2020)

This letter describes a problem that a Chinese university student like yourself is having. Please read the letter and decide what this student should do. You will have 2.5 min to explain your advice in English. You should do the following four things in order: (1) explain the problem, (2) compare possible solutions, (3) recommend one of these, and (4) give the reason for your recommendation.

References

- Ahmadian, Mohammad Javad. 2011. The Effect of ‘Massed’ Task Repetitions on Complexity, Accuracy and Fluency: Does It Transfer to a New Task? *Language Learning Journal* 39: 269–80. [CrossRef]
- Allard, Fran, and Janet L. Starkes. 1991. Motor-Skill Experts in Sports, Dance, and Other Domains. In *Toward a General Theory of Expertise*. Edited by Karl Andersson and Jackqui Smith. Cambridge: Cambridge University Press, pp. 126–52.
- Anderson, John R. 2009. *How Can the Human Mind Occur in the Physical Universe?* Oxford: Oxford University Press. ISBN 978-0-19-974126-7.
- Bao, Gui. 2023. Factor Structures of Speed and Breakdown Fluency in EFL Learners’ Story Retelling Performances. *International Review of Applied Linguistics in Language Teaching* 61: 631–54. [CrossRef]
- Boersma, Paul, and David Weenink. 2016. Praat: Doing Phonetics by Computer. Available online: <http://www.praat.org/> (accessed on 25 December 2016).
- Brunmair, Matthias, and Tobias Richter. 2019. Similarity Matters: A Meta-Analysis of Interleaved Learning and Its Moderators. *Psychological Bulletin* 145: 1029–52. [CrossRef]
- Bui, Gavin, Mohammad Javad Ahmadian, and Ann-Marie Hunter. 2019. Spacing Effects on Repeated L2 Task Performance. *System* 81: 1–13. [CrossRef]
- Bui, Gavin, Peter Skehan, and Zhan Wang. 2018. Task Condition Effects on Advanced-Level Foreign Language Performance. In *The Handbook of Advanced Proficiency in Second Language Acquisition*. Edited by Paul A. Malovrh and Alessandro G. Benati. New York: Wiley, pp. 219–37. ISBN 978-1-119-26161-2.
- Bygate, Martin. 2001. Effects of Task Repetition on the Structure and Control of Oral Language. In *Researching Pedagogic Tasks: Second Language Learning, Teaching and Testing*. Edited by Martin Bygate, Peter Skehan and Merrill Swain. Harlow: Pearson Education, pp. 23–48.
- Carvalho, Paulo F., and Robert L. Goldstone. 2017. The Sequence of Study Changes What Information Is Attended to, Encoded, and Remembered during Category Learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 43: 1699–719. [CrossRef] [PubMed]
- Cho, Minyoung. 2018. Task Complexity and Modality: Exploring Learners’ Experience From the Perspective of Flow. *The Modern Language Journal* 102: 162–80. [CrossRef]
- de Jong, Nel, and Charles A. Perfetti. 2011. Fluency Training in the ESL Classroom: An Experimental Study of Fluency Development and Proceduralization. *Language Learning* 61: 533–68. [CrossRef]
- de Jong, Nel, and Philip Tillman. 2018. Grammatical Structures and Oral Fluency in Immediate Task Repetition: Trigrams across Repeated Performances. In *Task-Based Language Teaching*. Edited by Martin Bygate. Amsterdam: John Benjamins Publishing Company, vol. 11, pp. 43–73. ISBN 978-90-272-0113-3.
- de Jong, Nivja H., and Hans R. Bosker. 2013. Choosing a Threshold for Silent Pauses to Measure Second Language Fluency. Paper presented at 6th Workshop on Disfluency in Spontaneous Speech (DiSS), Stockholm, Sweden, August 21–23; pp. 17–20.
- de Jong, Nivja H., and Ton Wempe. 2009. Praat Script to Detect Syllable Nuclei and Measure Speech Rate Automatically. *Behavior Research Methods* 41: 385–90. [CrossRef]
- DeKeyser, Robert. 2015. Skill Acquisition Theory. In *Theories in Second Language Acquisition. An Introduction*. Edited by Bill VanPatten and Jessica Williams. London: Routledge, pp. 94–112.
- Dörnyei, Zoltán, and Judit Kormos. 1998. Problem-solving mechanisms in L2 communication: A Psycholinguistic Perspective. *Studies in Second Language Acquisition* 20: 349–85. [CrossRef]
- Fillmore, Charles J. 1979. On Fluency. In *Individual Differences in Language Ability and Language Behavior*. Edited by Charles J. Fillmore, Daniel Kempler and William S.-Y. Wang. Cambridge, MA: Academic Press, pp. 85–101. ISBN 978-0-12-255950-1.
- Gerbier, Emilie, and Thomas C. Toppino. 2015. The Effect of Distributed Practice: Neuroscience, Cognition, and Education. *Trends in Neuroscience and Education* 4: 49–59. [CrossRef]
- Hunter, Ann-Marie. 2017. Fluency Development in the ESL Classroom: The Impact of Immediate Task Repetition and Procedural Repetition on Learners’ Oral Fluency. Ph.D. thesis, St Mary’s University, Twickenham, UK.
- Kang, Sean H. K., and Harold Pashler. 2012. Learning Painting Styles: Spacing Is Advantageous When It Promotes Discriminative Contrast: Spacing Promotes Contrast. *Applied Cognitive Psychology* 26: 97–103. [CrossRef]
- Kormos, Judit. 2006. *Speech Production and Second Language Acquisition*. Cognitive Sciences and Second Language Acquisition. Mahwah: Lawrence Erlbaum Associates. ISBN 978-0-8058-5657-6.
- Lambert, Craig, Judit Kormos, and Danny Minn. 2017. Task Repetition and Second Language Speech Processing. *Studies in Second Language Acquisition* 39: 167–96. [CrossRef]
- Lambert, Craig, Scott Aubrey, and Paul Leeming. 2020. Task Preparation and Second Language Speech Production. *TESOL Quarterly* 55: 331–65. [CrossRef]
- Lightbown, Patsy Martin. 2007. Transfer Appropriate Processing as a Model for Classroom Second Language Acquisition. In *Understanding Second Language Process*. Edited by ZhaoHong Han. Clevedon: Multilingual Matters, pp. 27–44. [CrossRef]
- Lightbown, Patsy Martin. 2019. Perfecting Practice. *The Modern Language Journal* 103: 703–12. [CrossRef]
- Nakata, Tatsuya, and Yuichi Suzuki. 2019. Mixing Grammar Exercises Facilitates Long-Term Retention: Effects of Blocking, Interleaving, and Increasing Practice. *The Modern Language Journal* 103: 629–47. [CrossRef]

- Pan, Steven C., Jahan Tajran, Jarrett Lovelett, Jessica Osuna, and Timothy C. Rickard. 2019. Does Interleaved Practice Enhance Foreign Language Learning? The Effects of Training Schedule on Spanish Verb Conjugation Skills. *Journal of Educational Psychology* 111: 1172–88. [\[CrossRef\]](#)
- Saito, Kazuya, Konstantinos Macmillan, Magdalena Kachlicka, Takuya Kuniyama, and Nobuaki Minematsu. 2022. Automated Assessment of Second Language Comprehensibility: Review, Training, Validation, and Generalization Studies. *Studies in Second Language Acquisition* 45: 234–63. [\[CrossRef\]](#)
- Segalowitz, Norman. 2010. *Cognitive Bases of Second Language Fluency*. New York: Routledge. ISBN 978-1-136-96883-9.
- Skehan, Peter. 2003. Task-Based Instruction. *Language Teaching* 36: 1–14. [\[CrossRef\]](#)
- Suzuki, Shungo, and Judit Kormos. 2023. The Multidimensionality of Second Language Oral Fluency: Interfacing Cognitive Fluency and Utterance Fluency. *Studies in Second Language Acquisition* 45: 38–64. [\[CrossRef\]](#)
- Suzuki, Yuichi. 2021a. Individual Differences in Memory Predict Changes in Breakdown and Repair Fluency but Not Speed Fluency: A Short-Term Fluency Training Intervention Study. *Applied Psycholinguistics* 42: 969–95. [\[CrossRef\]](#)
- Suzuki, Yuichi. 2021b. Optimizing Fluency Training for Speaking Skills Transfer: Comparing the Effects of Blocked and Interleaved Task Repetition. *Language Learning* 71: 285–325. [\[CrossRef\]](#)
- Suzuki, Yuichi. 2022. Automatization and Practice. In *The Routledge Handbook of Second Language Acquisition and Psycholinguistics*. New York: Routledge, pp. 308–21. ISBN 978-1-00-301887-2.
- Suzuki, Yuichi, Masaki Eguchi, and Nel De Jong. 2022a. Does the Reuse of Constructions Promote Fluency Development in Task Repetition? A Usage-Based Perspective. *TESOL Quarterly* 56: 1290–319. [\[CrossRef\]](#)
- Suzuki, Yuichi, Satoko Yokosawa, and David Aline. 2022b. The Role of Working Memory in Blocked and Interleaved Grammar Practice: Proceduralization of L2 Syntax. *Language Teaching Research* 26: 671–95. [\[CrossRef\]](#)
- Suzuki, Yuichi, Tatsuya Nakata, and Robert Dekeyser. 2019. The Desirable Difficulty Framework as a Theoretical Foundation for Optimizing and Researching Second Language Practice. *The Modern Language Journal* 103: 713–20. [\[CrossRef\]](#)
- Tavakoli, Parvaneh. 2009. Investigating Task Difficulty: Learners' and Teachers' Perceptions. *International Journal of Applied Linguistics* 19: 1–25. [\[CrossRef\]](#)
- Tavakoli, Parvaneh, and Clare Wright. 2020. *Second Language Speech Fluency: From Research to Practice*. Cambridge: Cambridge University Press. ISBN 978-1-108-49961-3.
- Tavakoli, Parvaneh, Colin Campbell, and Joan McCormack. 2016. Development of Speech Fluency Over a Short Period of Time: Effects of Pedagogic Intervention. *TESOL Q* 50: 447–71. [\[CrossRef\]](#)
- Thai, Chau, and Frank Boers. 2016. Repeating a Monologue Under Increasing Time Pressure: Effects on Fluency, Complexity, and Accuracy. *TESOL Quarterly* 50: 369–93. [\[CrossRef\]](#)
- Tran, Mai Ngoc, and Kazuya Saito. 2021. Effects of the 4/3/2 Activity Revisited: Extending Boers (2014) and Thai & Boers (2016). *Language Teaching Research*. [\[CrossRef\]](#)
- Zimmerman, Joel. 1975. Free Recall after Self-Paced Study: A Test of the Attention Explanation of the Spacing Effect. *The American Journal of Psychology* 88: 277. [\[CrossRef\]](#)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.