

ORIGINAL ARTICLE

Individual differences in memory predict changes in breakdown and repair fluency but not speed fluency: A short-term fluency training intervention study

Yuichi Suzuki* 

Department of Cross-Cultural Studies, Faculty of Cross-Cultural and Japanese Studies, Kanagawa University, Kanagawa, Japan

*Corresponding author. Email: szky819@kanagawa-u.ac.jp

(Received 28 August 2020; revised 4 April 2021; accepted 15 April 2021; first published online 11 May 2021)

Abstract

The purpose of this intervention study is to reveal the extent to which memory-related aptitudes are implicated in the second language (L2) fluency development fostered by task repetition. English L2 learners are engaged in oral narrative tasks three times per day under two different 3-day task repetition schedules: blocked (Day 1: A-A-A, Day 2: B-B-B, Day 3: C-C-C) versus interleaved (Day 1: A-B-C, Day 2: A-B-C, Day 3: A-B-C). Their phonological short-term memory (PSTM), attention control, and associative memory were used as predictors of fluency changes measured through speed, breakdown, and repair fluency behaviors. Results showed that while the articulation rate change was not explained by any of the examined predictors, breakdown and repair fluency were predicted by different memory components. Specifically, PSTM was associated with mid-clause pause decrease during the training phase, while associative memory was linked to the increase in clause-final pauses in the posttest. Attention control, as well as PSTM, was related to greater repair frequency in the posttest, indicating increased learners' attention to speech monitoring. Furthermore, PSTM and associative memory contributed to reducing breakdown fluency in the blocked repetition condition only, suggesting that learners can capitalize on their memory for improving oral fluency by engaging in blocked practice.

Keywords: individual differences; memory; fluency; task repetition; blocked and interleaved practice; aptitude–treatment interaction

A growing body of research suggests that cognitive aptitude, that is, an array of perceptual and cognitive abilities that are utilized for the second language (L2) learning, is an important explanatory factor for L2 attainment in classroom (e.g. Muñoz, 2014; Ranta, 2002; Saito, 2017, 2019), during study abroad (e.g. Grey et al., 2015; Tare et al., 2018), and in naturalistic contexts (e.g. Dąbrowska, 2019; Granena,

© The Author(s), 2021. 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 in any medium, provided the original work is properly cited.

2013; Suzuki & DeKeyser, 2017b). While individual differences in cognitive aptitudes can account for the variance in outcomes in specific linguistic domains, such as vocabulary and grammar acquisition (e.g. Li, 2016), the role of cognitive aptitudes in other aspects of L2 learning, such as speaking skills, is insufficiently explored.

However, findings yielded by the emerging line of aptitude research suggest that cognitive aptitudes influence L2 speaking ability, the fluency dimension in particular. In this context, fluency is defined as “flow, continuity, automaticity or smoothness of speech” (Koponen & Riggensbach, 2000, p. 6). The focus of the current study is utterance fluency that is typically broken down into speed fluency (e.g. articulation rate), breakdown fluency (e.g. pause frequency and duration), and repair fluency (e.g. self-repetition and reformulation) (Housen & Kuiken, 2009; Tavakoli & Skehan, 2005). Empirical evidence indicates the presence of potential links between some aspects of utterance fluency measured at one point in time and cognitive aptitudes such as phonological short-term memory (PSTM) (Granena & Yilmaz, 2019; Wen, 2016), working memory (Brown Nielson & DeKeyser, 2019; Roger & Carmen, 2010; Wen, 2016), and associative memory (Saito, 2017). Although the cross-sectional approach taken by these studies can reveal an association between cognitive aptitudes and L2 fluency performance at one specific point in time, it is difficult to establish a more direct connection between cognitive aptitude and fluency *development* (cf. O'Brien et al., 2007). In order to elucidate the role of aptitude in fluency development in more detail, the effects of aptitude should be examined using an intervention research design that fosters L2 fluency development.

One intervention technique that can be adopted to enhance L2 fluency development is task repetition, that is, “the repeated performance of tasks that share some of the same pragmatic purpose or purposes and some of the same content” (Bygate, 2018, p. 13). When L2 learners are allowed to repeat the same tasks (e.g. narrating the same story more than once), as relevant aspects of schematic and content memory are activated and remembered during the first task, more cognitive resources can be subsequently used for allocating attention to formulating linguistic forms accurately (Fukuta, 2016). The subsequent task performance is further enhanced by the lexical and syntactic encoding processes activated during the first task performance, resulting in fluency development. While task repetition has been found to facilitate fluency development (e.g. Ahmadian & Tavakoli, 2011; Bygate, 1996, 2001; de Jong & Perfetti, 2011; Lambert et al., 2017; Lynch & Maclean, 2000; Thai & Boers, 2016), some variations in the outcome of task repetition fluency training are observed. It is thus worth exploring the extent to which individual differences in cognitive aptitudes can account for fluency development fostered by task repetition.

Furthermore, in recent years, there is a surge of interest in investigating how individual differences in aptitude interact with instructional treatment in L2 research (DeKeyser, 2019; Vatz et al., 2013). This novel line of investigation is grounded in an educational psychology research paradigm denoted as aptitude–treatment interaction (ATI) owing most notably to the pioneering work of Cronbach and Snow (1977). The number of studies exploring ATI in the context of instructed L2 is growing consistently because understanding facilitative and inhibitory roles of aptitudes in L2 acquisition under specific learning conditions can shed light on the different mental processes induced by a particular learning condition (DeKeyser, 2012). Furthermore, ATI patterns can eventually be utilized to

individualize different practice activities based on learners' aptitude strengths and weaknesses for optimizing L2 learning (e.g. Erlam, 2005; Suzuki & DeKeyser, 2017a; Yilmaz, 2013). Because L2 speaking process is cognitively demanding requiring the execution of multiple processing components (Kormos, 2006; Skehan, 2009), capitalizing on learners' capacity to process L2 speech efficiently may be a promising step toward devising effective L2 speaking training. To the best of the author's knowledge, the role of cognitive aptitude in task repetition fluency training has never been investigated, due to which no data on the effectiveness of different types of task repetition presently exists. In a short-term intervention study, Suzuki (2020) compared blocked task repetition (whereby participants performed the same task multiple times within a single training session) and interleaved task repetition (participants were given different tasks within a single training session) on fluency development. While the findings of this study suggest that blocked task repetition yielded some advantages over interleaved task repetition, there were individual differences in fluency development among the participants assigned to both groups. Therefore, the aim of the present investigation is to reanalyze the data collected by Suzuki (2020) and uncover the role of cognitive aptitudes pertaining to memory – PSTM, attention control, and associative memory – in L2 fluency development through task repetition. Furthermore, potential ATI patterns are explored to examine the role of aptitude in fluency development under two different task repetition conditions.

The relationships between cognitive aptitude and L2 fluency

Cognitive aptitude is a multifaceted construct consisting of various perceptual and cognitive abilities that predict success in L2 learning in a variety of contexts. Memory is widely recognized as one of the central components of cognitive aptitude. The classic aptitude test battery, Modern Language Aptitude Test (MLAT), includes instruments assessing rote or associative memory ability, as well as three other components (phonetic coding ability, grammatical sensitivity, and inductive language learning ability). However, in a recent reconceptualization and extension of cognitive aptitude, which is informed by research findings from psychology, multiple memory functions are highlighted (Linck et al., 2013; Wen et al., 2017). Most notably, a comprehensive aptitude test battery, high-level language aptitude test battery (Hi-LAB), includes aptitude tests pertaining not only to long-term memory but also executive functioning in working memory (updating, inhibitory control, task switching subprocesses) as well as PSTM. Understanding the role of different facets of memory in L2 fluency development contributes to our understanding of cognitive processes underlying L2 speech fluency development. Because speaking an L2 requires a considerable amount of attentional resources, higher memory functioning should play an important role in L2 speech (Segalowitz, 2010).

Three aspects of memory-related aptitude components – PSTM, attention control, and associative memory – are highlighted in the current study on fluency training intervention. In extant cross-sectional studies, these memory components are found to be related to aspects of utterance fluency measured at one point in time. First, PSTM is a short-term storage component involving the rehearsal of

phonological information in L2 speech (Baddeley, 2012). In a cross-sectional study conducted by Granena and Yilmaz (2019), PSTM, measured by the letter span task, was linked to speed fluency (i.e. articulation rate) among L2 intermediate Spanish classroom learners.

Second, attention control is governed by the central executive system of working memory and concerns the capacity to control attentional resources during L2 speech (Engle, 2002). Yet, there is a paucity of research exploring the relationship between attention control and utterance fluency. In one of the few available studies, Zuniga and Simard (2019) focused on repair fluency in L2 speech and examined the link between attention control and L2 self-repairs in L2 English learners with French L1 who grew up in a French-speaking household in Quebec. These authors found that learners with higher attention control, measured by trail-making task (TMT), tended to make fewer repairs. This finding suggests that learners exhibiting a higher capacity to regulate their attentional resources may be more likely to avoid making self-repairs during L2 speech.

Third, associative memory in the long-term memory system – the ability to remember the relations among multiple items – was also found to be related to L2 fluency. Saito (2017) focused his investigation on L2 English learners recruited from a Japanese university and examined how speed fluency was predicted by four components of LLAMA aptitude tests. The findings indicated that, among the four subcomponents, LLAMA_B subtest (which measures associative memory ability) was the only significant predictor of speed fluency (i.e. articulation rate), but not pause ratios. Although Saito (2017) did not distinguish the location of pauses, analyzing different aspects of breakdown fluency is useful, because mid-clause and clause-final pauses can be linked to different cognitive processes involved in speaking (de Jong, 2016; Kahng, 2018; Lambert et al., 2017; Skehan et al., 2016), which may draw on different aptitude components.

In sum, previous research suggests that faster articulation rate, which is one aspect of speed fluency, is likely linked to higher PSTM and associative memory, while higher attention control is associated with fewer repairs. Although these associations between aptitude and fluency suggest that memory potentially plays a facilitative role in fluency development, two limitations must be noted when interpreting the aforementioned findings. First, the scope of prior L2 fluency studies was limited in the number of memory components as well as utterance fluency measures examined. In fact, none of the extant studies involved more than one aspect of memory, which was examined for a handful of fluency measures without covering all three fluency dimensions of speed, breakdown, and repair fluency. It is thus important to examine the roles of multiple cognitive aptitude components in different aspects of fluency development, which sheds light on L2 cognitive processes underlying fluency development. One theory of the development of L2 speech processing capacity (e.g. Kormos, 2006; Skehan, 2009) stipulates that lower proficiency L2 learners often need to engage in serial processing in which content message and linguistic encoding occur sequentially. As L2 proficiency increases, L2 learners gradually shifts to more efficient, parallel processing of content and language. The shift of serial to parallel processing can be potentially captured by analyzing pauses that occur within the clausal boundary because they largely reflect linguistic encoding, as opposed to clause-final pauses primarily indicating conceptualization (Lambert et al., 2017,

2020). Moreover, Skehan (2016) proposed that fluency development can be linked particularly to a putative L2 learning stage pertaining to gaining control over L2 knowledge. In this later stage of L2 learning, the existing knowledge can be accessed more quickly or proceduralized to support fluent real-time L2 use, such as during speech production. Hence, establishing the relationships between memory components (e.g. PSTM, attention control, and associative memory) and some aspects of fluency development (e.g. speed fluency and mid-clause pauses) would allow identification of memory components that contribute to more efficient L2 speech processing.

Second, the authors of all the studies reviewed here adopted a cross-sectional design. While O'Brien et al. (2007) found that PSTM measured by the nonword serial recognition task predicted the long-term oral fluency development (e.g. speech rate, mean run length, filled pauses) over 13 weeks in the regular classroom and study abroad contexts, the role of memory in fluency development catered by short-term intensive fluency training intervention has not yet been examined. Given the paucity of research in this domain, this gap is addressed in the current study based on a task repetition intervention design. In particular, as suggested by Skehan (2016), task repetition not only enhances proceduralization of existing knowledge, but also compensates for the learner's weakness in some cognitive aptitudes through engagement in repeated practice. It is thus worth exploring whether fluency development fostered by task repetition is influenced by individual differences in cognitive aptitudes.

Aptitude–treatment interaction in L2 learning

Findings yielded by an increasing number of L2 studies focusing on ATI indicate that learners' cognitive aptitudes moderate the effectiveness of different types of L2 instruction (see DeKeyser, 2019; Vatz et al., 2013 for an overview). A vast majority of empirical ATI studies examined L2 grammar acquisition under different instructional/learning conditions such as incidental versus intentional learning (e.g. Robinson, 1997), inductive versus deductive teaching (e.g. Erlam, 2005; Hwu et al., 2014), or explicit versus implicit corrective feedback (e.g. Li, 2013; Yilmaz, 2013).

One phenomenon that has recently started to attract increased attention is the role of cognitive aptitudes in different L2 grammar practice schedules. For instance, Suzuki and colleagues revealed that while language analytic ability and associative memory facilitate L2 grammar acquisition under longer spaced learning condition (e.g. 7-day interval), working memory may be more related to L2 learning under shorter spaced condition (e.g. 1-day interval) (Suzuki, 2019; Suzuki & DeKeyser, 2017a). The most relevant study to the current investigation was reported by Suzuki, Yokosawa, and Aline (2020), who examined the role of working memory capacity in the context of blocked and interleaved practice aimed at L2 grammar acquisition. In their study, English L2 learners, all of whom were recruited from a Japanese university, performed a simple oral picture description task that specifically elicited a relative clause construction. For instance, participants were presented with a picture depicting a boy kissing a dog, and were instructed to describe it using

a relative clause (e.g. “That is the boy who is kissing the dog.”). The learners engaged in these narrowly focused grammar practice tasks targeting five relative clause constructions (who, whom, which [subject and object], and where) under either blocked or interleaved practice conditions. Specifically, in the blocked practice condition, the practice items were blocked by grammatical category (e.g. 10 instances involving *who*, followed by 10 instances involving *whom*), whereas in the interleaved practice condition, practice items from different categories were intermixed (e.g. who, which, whom, where, who, which, etc.). The study findings revealed an ATI pattern indicating that working memory played a facilitative role in accelerating articulation rate of relative clause sentence in the blocked group only (a similar pattern was also found by Sana et al. (2017), who focused on concept learning).

While Suzuki et al. (2020) revealed that the effects of blocked practice were contingent on individuals’ cognitive ability, the authors mainly focused on the acquisition of one aspect of grammar knowledge (i.e. relative clause construction) gained through narrowly focused grammar practice. Extending the scope of this prior investigation to more meaning-focused practice with the aim of developing utterance fluency, the current study focuses on blocked and interleaved task repetition fluency training using picture narratives. For this purpose, the data gathered by Suzuki (2020) as a part of an earlier investigation, wherein English L2 learners in a Japanese university engaged in task repetition fluency training for 3 consecutive days outside of the classroom, were reanalyzed from ATI perspectives. The participants in this prior study performed three oral narrative tasks every day under either blocked (Day 1: AAA, Day 2: BBB, Day 3: CCC) or interleaved (Day 1: ABC, Day 2: ABC, Day 3: ABC) task repetition condition. Changes in their performance from pretest to posttest were analyzed using novel oral narrative tasks (i.e. different from those used in the training sessions). Although the results showed that blocked task repetition led to greater fluency development than interleaved task repetition, under both task repetition conditions, there were considerable variations in the rate of fluency progress both during the training and from the pretest to the posttest. The current investigation focuses on these individual differences in fluency development, aiming to link them to different components of memory abilities to elucidate cognitive processes underlying speech fluency under blocked and interleaved task repetition conditions.

The current study and research questions

By reanalyzing the fluency data obtained as a part of the intensive fluency training intervention conducted by Suzuki (2020) in relation to learners’ aptitudes, this study aims at investigating the extent to which memory-related aptitudes are recruited for different aspects of L2 fluency development. In the current study, the roles of three memory components – PSTM, attention control, and associative memory – were examined simultaneously in fluency development in terms of the three dimensions (speed, breakdown, and repair fluency). Furthermore, the ATI analyses allowed us to explore whether the relationships between these memory components and fluency measures change by the task repetition intervention type – blocked or

interleaved task repetition – both during the three training sessions and in the pretest–posttest changes. Two research questions were addressed:

1. To what extent do individual differences in memory abilities (PSTM, attention control, and associative memory) predict the development in three fluency facets (speed, breakdown, and repair) during training and in the pretest–posttest change?
2. Are the effects of memory abilities moderated by the task practice sequence (blocked and interleaved task repetition practice)?

Methods

Participants

The sample for this study comprised of 50 L2 English learners attending a Japanese university (aged 18–22) who were recruited through announcements in their regular EFL classes. Their English proficiency was estimated to fall between A2 (elementary) and B1 (intermediate) levels on the Common European Framework of Reference for Languages (CEFR) benchmark. They were randomly assigned to either a blocked task repetition ($n = 24$) or an interleaved task repetition ($n = 26$) condition. The difference between the English proficiency levels of students assigned to the two groups – based on the junior English Minimal Test, an objective proficiency test developed for research purposes (Goto et al., 2010) – was not statistically significant.

Aptitude tests

Nonword repetition task

The nonword repetition task (NWRT) for Japanese speakers, which was developed and validated by Yamaguchi and Shimizu (2011), was used as a measure of PSTM. Nonwords (three-mora nonwords generated by combining three Japanese morae randomly) were presented to the participants through headphones, and they were required to repeat them accurately. This NWRT utilized L1 phonology in order to assess PSTM independently from L2 English proficiency. The test consisted of 20 trials, with 5 trials for each set consisting of 1, 2, 3, or 4 nonwords. For instance, in a two nonword trial, participants heard and repeated two nonwords (e.g. *teafu*, *hetesu*). Repetition attempts were recorded during the testing sessions, and each word was scored by a Japanese native speaker. The total number of successful repetitions (with 50 as the maximum) was used as a PSTM score, indicating how well participants store phonological information and articulate the words. Data were not obtained from one participant due to a technical error in recording. The mean score (SD) based on the remaining sample was 28.43 (5.89) and the Cronbach's alpha was .82.

Trail making task

The TMT was chosen as a measure of attention control in this study, as it is not only used for clinical purposes (e.g. identifying patients with attentional deficits) but also

for L2 research (Trofimovich et al., 2007; Zuniga & Simard, 2019). As attention and memory are different but related constructs responsible for executive functioning in working memory (Miyake & Friedman, 2012), TMT should not be seen as a purely memory measure, but a measure of attention control. The current TMT was a Japanese translation version of the Halstead-Reitan Battery subtest (Reitan & Wolfson, 1985) that was validated by Japanese researchers (Abe et al., 2004). The TMT consists of two parts. Part A is a baseline test measuring general information processing speed. Using a pencil, participants needed to connect the numbers from 1 to 25 that were randomly scattered on the test sheet as quickly as possible. Part B is a more complex test that measures how efficiently participants can switch their attention between two types of information. The test sheet for Part B included both numbers (1–12) and a sequence of 13 Japanese *kana* characters (from あ to し). Participants had to connect the alternating numbers and characters (i.e. 1-あ-2-い-3-う-4-え), which required repeatedly shifting attentional resources from numbers to letters to generate the correct sequence. The time participants took to complete both parts was measured using a stopwatch. The difference score (Part B–Part A) was used to index attention control ability, indicating how efficiently participants could allocate their attention in the simultaneous task (Part B) against the baseline processing speed (Part A). Data pertaining to three participants were excluded from analyses because they failed to correctly connect the numbers and characters. The mean score (Part B–Part A) based on the remaining sample was 29.75 ($SD = 18.04$). In order to facilitate the interpretation of results yielded by correlational and regression analyses, an inverse score (i.e. Part A–Part B) was used to align the direction of the scores to those related to the other two aptitude tests (i.e. higher scores indicate greater memory abilities).

LLAMA_B

LLAMA_B (Meara, 2005) was used to measure participants' associative memory ability. The test consists of a learning and a test session. In the learning phase, participants were presented with a panel on the computer screen consisting of 20 animated pictures that were associated with novel words. They had to remember as many words as possible within 2 min. In the subsequent test phase, they were presented with each of the 20 words in written form and were instructed to choose a corresponding picture on the computer. *LLAMA_B* is thus an associative memory task that is relatively independent of a particular language. The total number of correct responses indicated how well participants could memorize the vocabulary. The mean score (SD) on this test was 13.60 (3.96) and the Cronbach's alpha was .79.

Materials

Training materials

Three prompts (Bicycle, Tiger, and Race) were used for oral narrative practice. These prompts were originally developed by Heaton (1996) and all three picture stories and the guiding questions were adopted from the previous research on L2 fluency (de Jong & Tillman, 2018; de Jong & Vercellotti, 2016). Each prompt comprised of six-panel picture stories based on a tight sequential structure with

	Week 1	Week 2				Week 3
		Day 1	Day 2	Day 3	Day 4	
Blocked	Pretest	AAA	BBB	CCC	Posttest	Aptitude tests
Interleaved		ABC	ABC	ABC		

Figure 1. Experimental Schedules.

a similar narrative structure (involving little causal reasoning) and the number of elements (e.g. main characters, locations).

Pretest and posttest

In order to assess fluency development, two different six-panel picture stories (Street and Airport) were used in the pretest and posttest. These two stories also had a tight sequential structure with a similar narrative structure involving little causal reasoning (i.e. a thief steals the main character's purse/suitcase, and another main character helps catching the thief). Each of the prompts involved three main characters (thief, victim, helper) at a different location (street and airport). The picture prompts used for pretest and posttest were counterbalanced. All instruments are available in the IRIS digital repository of data collection instruments (Marsden et al., 2016).

Procedure

As shown in Figure 1, a pretest – training – posttest design was used in the present study. One week prior to the training session, participants took a pretest in a computer lab. After they were randomly assigned to one of the experimental conditions (blocked or interleaved practice), they engaged in a 3-day fluency training program outside the lab (Days 1, 2, and 3) by following the instructions and recorded their narratives using an IC recorder. Participants assigned to the blocked task repetition condition performed the same narrative task three times a day (e.g. AAA-BBB-CCC), whereas those in the interleaved task repetition condition performed three different narrative tasks on each of the 3 days (e.g. ABC-ABC-ABC). The order of the three prompts (i.e. Bicycle, Tiger, and Race) was counterbalanced for each participant. The participants were allowed the same amount of time (i.e. 3 min) for each narration throughout the experiment. To ensure that participants performed the fluency training as indicated, a research assistant sent a daily reminder to them. On the day following the last training session (Day 4), participants came to the computer lab for the posttest and also took the aptitude tests later during the subsequent week. Because aptitude may improve with extensive linguistic experience (Huang et al., 2020), the aptitude measures should have been taken prior to the experimental treatment (e.g. before Week 1). This was not logistically feasible, however. Nonetheless, given that the current 3-day fluency training duration was short, the influence of the intervention presumably had little effect on the participants' aptitude test scores.

Table 1. Speed, breakdown, and repair fluency measures

Category	Measure	Operationalization
Speed fluency	Articulation rate	Number of syllables per minute of speech, excluding pauses
Breakdown fluency	Mid-clause pause duration	Mean duration of mid-clause filled and unfilled pauses
	Clause-final pause duration	Mean duration of clause-final filled and unfilled pauses
	Mid-clause pause frequency	Number of mid-clause filled and unfilled pauses per minute
	Clause-final pause frequency	Number of clause-final filled and unfilled pauses per minute
Repair fluency	Repetition frequency	Number of self-repetitions per minute
	Repair frequency	Number of reformulations and replacements per minute

Data coding

A total of 550 speech datasets, derived from pretest (50 learners), training (50 learners \times 9 deliveries), and posttest (50 learners), were coded by 3 trained coders. As shown in Table 1, seven fluency measures were computed for speed fluency (articulation rate), breakdown fluency (mean length of mid-clause and clause-final pauses and frequency of mid-clause and clause-final pauses per minute), and repair fluency (frequency of repetition and repairs per minute).

These seven measures can capture three dimensions of utterance fluency (Housen & Kuiken, 2009; Skehan, 2009). First, articulation rate was chosen as a measure of speed fluency. Note that in the original data analyses conducted by Suzuki (2020), two additional speed fluency measures – mean length of run (number of syllables between pauses including false starts and repetition but excluding fillers) and phonation/time ratio (utterance duration divided by the total duration) – were also computed. However, these were excluded from the current analyses because they are composite measures of fluency speed and breakdown, and the interpretations of their association with the aptitude component were not as straightforward as those related to the other purer fluency measures examined in this study. Second, for the four measures of breakdown fluency, pauses were defined as the filled and unfilled (silent) pauses lasting at least 200 ms (de Jong & Perfetti, 2011) and were further coded for mid-clause pauses (i.e. within the Analysis of Speech [AS] unit, Foster et al., 2000) and clause-final pauses (i.e. at the boundary of the AS unit). Mid-clause pauses indicate linguistic breakdown such as lexical and syntactic ones, whereas clause-final pauses presumably reflect conceptualization, including content planning (de Jong, 2016; Kahng, 2018; Lambert et al., 2017, 2020; Skehan et al., 2016). Third, repair fluency (the repetition and repair frequency) is often subsumed under utterance fluency, but the constructs are slightly different from the other two. In particular, repair frequency refers to both overt reformulations (e.g. the poster which caution the . . . which tell us the tiger is so danger) and

replacements (e.g. feel sleep...sleepy; owner tell...told) in the current study. These self-repair behaviors are considered to reflect learners' attentional resources directed to monitoring their speech and reformulating initially encoded language (Kormos, 1999; Lambert et al., 2017, 2020). Relatedly, repair frequency may also be considered as a measure of "attempted accuracy" because it indicates learners' self-directed attention to form and their effort to produce more accurate speech (Brown Nielson & DeKeyser, 2019; Gilabert, 2007). In contrast, repetition frequency simply refers to different dysfluency phenomena involving self-repetition (e.g. the man hit...hit the tiger's head) potentially indicating a coping strategy to buy time to linguistic encoding.

Statistical analysis

Hierarchical multiple regression analyses were conducted on the 3-day training gains (obtained by subtracting the score on the third performance from the one on the first performance for each day) and the pretest–posttest gain (the difference between the pretest and posttest score) for each of the seven fluency measurements. The hierarchical regression analyses consisted of three steps (Keith, 2015). First, Group (blocked vs. interleaved) was entered into the model as a predictor of fluency score change (Step 1). Next, three aptitude test scores were entered as predictors (Step 2). Finally, two-way interactions between Group (blocked vs. interleaved) and each aptitude test score were added (Step 3). If the interaction term in Step 3 was statistically significant, it was visualized using the "Interaction" program, version 1.7.2211 (Soper, 2006). The main focus of the current study concerned analyses conducted as a part of Step 2 and Step 3, and they, respectively, address the first research question (predictors of fluency development) and the second research question (ATI). Outliers were identified for each fluency measure ($z > 3.29$, Tabachnick & Fidell, 2013) and were excluded from the analysis (see Appendix S1 in the Online Supplementary File). Histograms of all fluency measures and three predictor variables indicated that the normality assumption was met for all variables. The assumption of multicollinearity was also met for all analyses, as VIF was less than 10 and tolerance was above .02 (Field, 2009).

Results

Descriptive statistics

As the descriptive statistics of fluency measures for both groups are presented in Appendix S2, only the major findings are summarized here. The blocked task repetition led to faster articulation, shorter mid-clause pause duration, and greater clause-final frequency than the interleaved task repetition with medium effect sizes both during training and in posttest relative to pretest. It is interesting to note that, in the blocked task repetition, which was found to be more effective (according to the articulation rate and mid-clause pause duration), clause-final frequency was *higher*. A similar pattern was also observed in the current ATI analysis – *increase* in clause-final frequency was associated with *higher* aptitudes (see below). Because the focus of the current study is restricted to the role of individual differences (rather

Table 2. Significant predictors of training performance and pretest–posttest changes

	Predictor		
	NWRT (PSTM)	TMT (Attention control)	LLAMA_B (Associative memory)
Day 1	Mid-clause pause dur. (↓)		Clause-final freq. (↓)
Day 2	Clause-final freq. (↑)		
Day 3	Mid-clause pause freq. (↓)		Mid-clause pause dur. (↓) Clause-final pause freq. (↑)
Pre-Post	Repairs (↑)	Repairs (↑)	Clause-final pause freq. (↑)

Note. Upward arrows indicate an increase in fluency measure corresponding to a higher aptitude score, whereas downward arrows indicate a decrease in fluency measure corresponding to a higher aptitude score. The correlation coefficients between the aptitude scores and fluency changes are provided in Appendix S3, and the full regression analysis results are presented in Appendix S4 in the Online Supplementary File.

than the main effects of blocked and interleaved task repetition) in fluency development, please refer to Suzuki (2020) for further details and discussion. The three predictors were not strongly correlated to each other (NWRT \times TMT: $r = .24$, $p = .11$; NWRT \times LLAMA_B: $r = .08$, $p = .60$; LLAMA_B \times TMT: $r = -.24$, $p = .10$), suggesting that the three aptitude tests tapped into different aspects of memory.

Predictors of fluency development (Research question 1)

Training performance data

The hierarchical regression analyses conducted as a part of Step 2 revealed significant predictors for fluency development during training (see Table 2). While none of the aptitude measures were significant predictors of four aspects of fluency measures (articulation rate, clause-final pause duration, repetition, and repairs), three fluency measures (mid-clause pause duration and frequency as well as clause-final pause frequency) were significantly predicted by the aptitude measures, which are delineated for each training day below.

Based on the data gathered on Day 1, two significant regression models were identified. The first model indicated that the addition of aptitude scores marginally significantly contributed to the further explanation of variance in mid-clause pause duration change, $\Delta F(3,38) = 2.59$, $\Delta R^2 = .17$, $p = .07$. According to this model, NWRT predicted the decrease in mid-clause pause duration ($B = -0.01$, $p = .01$). Second, although the addition of aptitude scores did not yield a significant improvement in the model in terms of explaining the clause-final frequency, $\Delta F(3,41) = 2.08$, $\Delta R^2 = .13$, $p = .12$, LLAMA_B score was a significant predictor of clause-final pause frequency change ($B = -0.18$, $p = .04$). When only LLAMA_B score was included in this model, it explained a similar amount of variance and the F change was significant, $\Delta F(1,46) = 5.94$, $\Delta R^2 = .11$, $p = .02$. This suggests that a higher LLAMA_B score was associated with a decrease in clause-final frequency on Day 1.

When applied to the data gathered on Day 2, only one model was significant. Specifically, the addition of aptitude scores significantly contributed to the further

explanation of variance in clause-final pause frequency change, $\Delta F(3,41) = 3.10$, $\Delta R^2 = .15$, $p = .04$. In this model, NWRT predicted the increase in clause-final pause frequency ($B = 0.14$, $p = .04$).

When data pertaining to Day 3 were analyzed, three significant models were found. First, the addition of aptitude scores contributed to the further explanation of variance in mid-clause pause duration, $\Delta F(3,41) = 2.31$, $\Delta R^2 = .11$, $p = .09$. Although this model was only marginally significant, LLAMA_B score was a significant predictor of mid-clause pause duration change ($B = -0.01$, $p = .02$), suggesting that a higher LLAMA_B score was associated with shorter mid-clause pause duration. Second, the addition of aptitude scores significantly contributed to the further explanation of variance in mid-clause pause frequency change, $\Delta F(3,41) = 3.15$, $\Delta R^2 = .19$, $p = .04$. In this model, NWRT predicted the decrease in mid-clause pause frequency ($B = -0.16$, $p = .04$). Last, although the addition of aptitude scores did not yield significant improvement in the model in terms of explaining the clause-final frequency, $\Delta F(3,41) = 1.98$, $\Delta R^2 = .10$, $p = .13$, LLAMA_B score was a significant predictor clause-final pause frequency change ($B = 0.19$, $p = .04$). Note that when only LLAMA_B was included in this model, the F change was significant and a similar amount of variance was explained solely by LLAMA_B, $\Delta F(1,46) = 6.17$, $\Delta R^2 = .09$, $p = .02$. This suggests that a higher LLAMA_B score was associated with an increase in clause-final frequency on Day 3.

Pretest–posttest data

Hierarchical regression analyses revealed significant predictors of pretest–posttest fluency development. As shown in Table 2, the aptitude played a significant role in the pretest–posttest changes in two out of seven fluency measures. First, the addition of aptitude scores was significant for predicting the number of repairs (Step 2), $\Delta F(3,41) = 3.05$, $\Delta R^2 = .18$, $p = .04$. In this model, NWRT and TMT predicted a significant increase in the repair frequency ($B = 0.09$, $p = .02$; $B = 0.03$, $p = .047$, respectively). Second, although the addition of aptitude scores failed to reach a significant level in the model aimed at explaining the clause-final frequency change, $\Delta F(3,41) = 2.12$, $\Delta R^2 = .13$, $p = .11$, LLAMA_B score was a significant predictor of clause-final pause frequency change ($B = 0.23$, $p = .048$). When only LLAMA_B was included in the model, it explained a similar amount of variance and the F change was significant, $\Delta F(1,46) = 4.75$, $\Delta R^2 = .09$, $p = .035$, suggesting that a higher LLAMA_B score was associated with an increase in clause-final frequency.

Aptitude–treatment interaction (Research question 2)

Training performance data

In Step 3 of the hierarchical multiple regression, interaction terms were added to the previous model (developed as a part of Step 2), and four statistical models yielded significant R^2 changes.¹ These four ATI patterns are illustrated graphically in Figure 2 (see Appendix S5 for the correlations between aptitude and fluency change in each group).

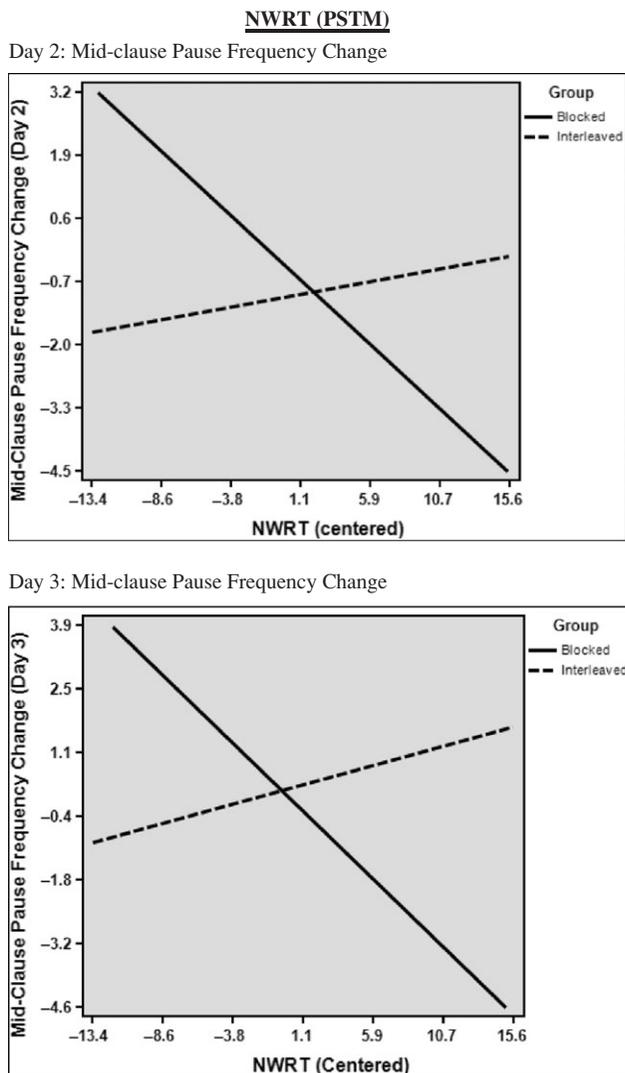
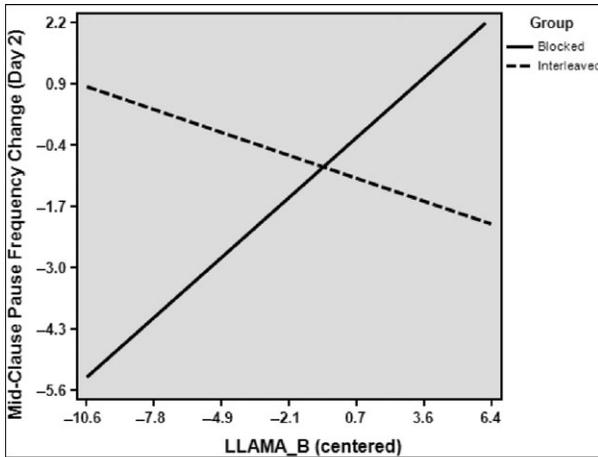


Figure 2. Significant Interactions Between Aptitude and Fluency Change During Training.

Two significant interactions between NWRT and Group were detected for the mid-clause pause frequency change on Day 2 ($B = 0.40$, $p = .04$) and Day 3 ($B = 0.42$, $p = .01$). As shown in Figure 2, higher NWRT score was related to a decrease in mid-clause pause frequency in the blocked practice group. A significant negative correlation was found between NWRT score and mid-clause pause frequency change during the training sessions on Day 2 ($r = -.46$, $p = .03$) and Day 3 ($r = -.63$, $p = .001$) in the blocked practice group only.

LLAMA_B (Associative Memory)

Day 2: Mid-clause Pause Frequency Change



TMT (Attention Control)

Day 3: Repetition Change

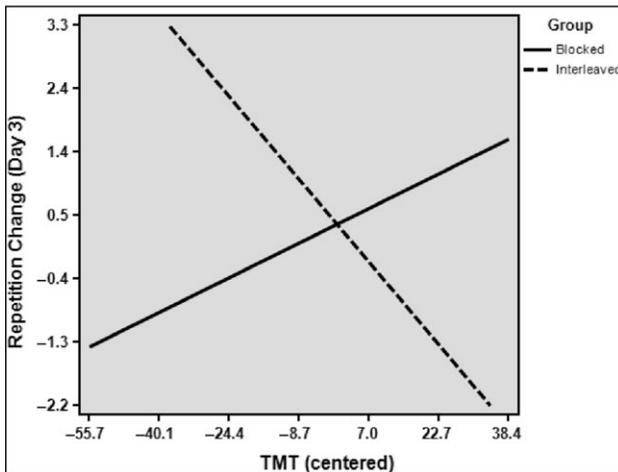


Figure 2. Continued
 Note. See Appendix S4 for the detailed statistical information.

The interaction between Group and LLAMA_B score was also significant ($B = -0.69, p = .01$). Interestingly, LLAMA_B score was positively correlated with the mid-clause pause frequency in the blocked group ($r = .47, p = .02$).

Furthermore, the interaction between Group and TMT was significant for the repetition change on Day 3 ($B = -0.11, p = .046$). This cross-over interaction suggests that the relationship between TMT and the repetition was positive in the blocked practice group, whereas the negative relationship was found in the interleaved practice group. The correlation between TMT and repetition change on

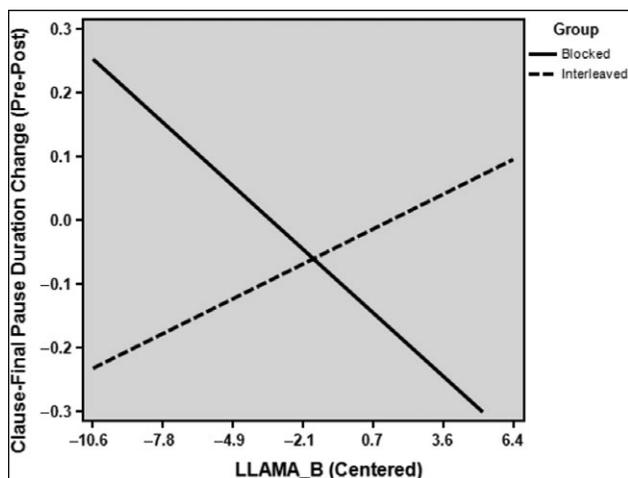


Figure 3. Significant Interactions Between Group and LLAMA_B Score for Pretest–Posttest Clause-Final Pause Duration Change.

Note. See Appendix S4 for the detailed statistical information.

Day 3 was slightly stronger in the interleaved group ($r = -.33, p = .12$) than in the blocked group ($r = .23, p = .30$).

Pretest–posttest data

When applied to the pretest–posttest changes, only one hierarchical multiple regression model produced significant R^2 changes following the addition of the interaction term (see Appendix S5 for the correlations between aptitude and fluency change in each group). A significant interaction between Group and LLAMA_B score was found for the clause-final pause duration changes ($B = 0.06, p = .02$). The cross-over ATI pattern in Figure 3 shows that higher LLAMA_B score contributed to shorter clause-final pause duration in the blocked group, whereas higher LLAMA_B score contributed to longer clause-final pause duration in the interleaved group. However, the correlation between clause-final pause duration and LLAMA_B score was significant in the blocked group ($r = -.44, p = .04$), but not in the interleaved group ($r = .30, p = .16$).

Discussion

Predictors of fluency development through task repetition

The first research question addressed in the present study concerned the roles of three memory components (PSTM, attention control, and associative memory) in fluency development both during the fluency training and the pretest–posttest changes. Three predictors failed to account for an aspect of speed fluency change (i.e. articulation rate) throughout the intervention period. However, multiple regression models incorporating the aptitude predictors (Step 2) explained 11–19% of variance in the breakdown fluency development (i.e. mid-clause pause frequency and

duration, as well as clause-final pause frequency) during the training sessions. When applied to the pretest–posttest changes, the significant model accounted for the additional 13% and 18% of the variance in the changes of one aspect of breakdown fluency (i.e. the clause-final pause frequency) and one aspect of repair fluency (i.e. repairs), respectively. Overall, the amount of variance explained by aptitude was consistent with the findings yielded by Li's (2016) meta-analysis on the relationship between aptitude and global L2 speaking skill attainment ($r = .37$, $R^2 = 13.6\%$).

In the remainder of this section, the findings related to the three aspects of fluency (speed, breakdown, and repair) examined in this work are discussed to elucidate cognitive processes underlying L2 fluency development. First, the lack of association between aptitude and speed fluency (i.e. articulation rate) is interpreted in light of the results obtained in previous cross-sectional studies. Second, given that PSTM was associated with the reduction in mid-clause pause behaviors, whereas associative memory was related to the changes in clause-final pause frequency, these distinct roles of the two memory components in different aspects of breakdown fluency are interpreted to elucidate the underlying L2 learning mechanisms involved in L2 speaking skills. Last, the positive contributions of PSTM and attention control to the increased number of repairs in the posttest relative to pretest are discussed.

Speed fluency: Limited role of memory in articulation rate during training and pretest–posttest changes

One surprising finding of this intervention study is the lack of systematic association between speed fluency (i.e. articulation rate) and any of the three aptitude components. This counters the findings yielded by previous cross-sectional studies indicating that articulation rate was related to PSTM (Granena & Yilmaz, 2019) and associative memory (Saito, 2017). This disparity could be due to the nature of the articulation rate measure. While articulation rate is associated with L2-specific skills such as vocabulary and grammar processing (de Jong et al., 2013), it is also largely influenced by L1 speaking styles (e.g. de Jong and Mora, 2019). Articulation may involve speech skills that are less susceptible to individuals' memory ability. In other words, L2 learners may be able to improve their articulation rate through task repetition regardless of the memory abilities examined in this study (cf., Skehan, 2016). Another possibility is that the current intervention was relatively short (i.e. nine task performances over 3 consecutive days), and improvements in the articulation rate might have reflected changes in narrower aspects of L2 fluency, such as faster lexical retrieval. As a case in point, when the effects of PSTM were examined in the context of long-term fluency development (i.e. L2 learning over 13 weeks both in classroom and study abroad settings), articulation rate acceleration was predicted by PSTM (O'Brien et al., 2007). Thus, it remains to be established to what extent cognitive aptitude contributes to speed fluency development in a more extended intervention study.

Breakdown fluency: Selective roles of PSTM and associative memory in mid-clause and clause-final pause changes

During the training phase, PSTM was consistently facilitative for reducing the mid-clause pause duration and frequency. Specifically, higher PSTM was associated with

shorter mid-clause pauses on Day 1 and with lower mid-clause pauses frequency on Day 3. These findings suggest that PSTM contributed to diminished breakdown fluency. As mid-clause pause presumably reflects a disruption in L2 linguistic encoding (e.g. lexical and grammatical) processes (Kahng, 2014), learners with higher PSTM would be more capable of counteracting such linguistic breakdown, which might have contributed to more efficient L2 linguistic processing. This observation extends the results reported by O'Brien et al. (2007), who found that PSTM played a significant role in the progress in breakdown fluency (i.e. mean length of run without filler or filled pauses) during 13-week classroom and study abroad experience, to the current short-term fluency training intervention. More broadly, from the perspective of speech processing model (Kormos, 2006; Lambert et al., 2017, 2020; Skehan, 2009), the present findings suggest that learners with higher PSTM were more likely to execute parallel processing of content and language during task repetition practice.

While the link between PSTM and mid-clause pause was evidenced in the present study, greater associative memory scores were consistently associated with *higher* clause-final frequency during the latter training phase² and in the posttest relative to pretest. Unlike mid-clause pause frequency, clause-final frequency is not necessarily a manifestation of speakers' disfluency, because it is made at an appropriate speech boundary (i.e. AS Unit). As posited by several authors, clause-final pause frequency is associated with conceptualization during L2 speech (de Jong, 2016; Kahng, 2018; Lambert et al., 2017; Skehan et al., 2016). Conceptualization is comprised of macroplanning and microplanning. While macroplanning (i.e. specification of what information one intends to convey) is relatively independent of L2-specific processing (Kormos, 2006), microplanning is closely related to L2-specific syntactic encoding because it refers to the establishment of a propositional form of the message (Levelt, 1989).

The fluency training provided to the current participants required limited cognitive demand related to macroplanning as the content of the speech was predefined by six frame pictures. Consequently, they could allocate most of their cognitive resources to microplanning involving L2 syntactic encoding. It can thus be speculated that greater associative memory capacity might have made it possible for learners to attempt to link their existing L2 knowledge with the propositions they were expected to express. In other words, the fluency intervention training design adopted in the present study might have allowed learners with higher associative memory to produce linguistic expressions that are more consistent with the unit of content planning (e.g. chunks rather than isolated single words), resulting in a greater number of pauses at the appropriate clause boundary. Recall that higher clause-final frequency was also associated with positive fluency development in the blocked task repetition group (see 4.1 Descriptive Statistics section as well as Suzuki, 2020), which also lends some support to the supposition that higher clause-final pause frequency may signify some developmental changes that can be assisted by superior associative memory. In sum, the selective roles of PSTM and associative memory revealed in this study suggest that mid-clause and clause-final pauses reflect different cognitive processes. As these roles were not completely exclusive (e.g. PSTM was also related to clause-final frequency pause on Day 2, and associative memory was related to mid-clause pause duration on Day 3), they warrant further investigation.

Table 3. Summary of ATI findings

	PSTM	Attention control	Associative memory
Day 1			
Day 2	Mid-clause pause freq. (Blocked [↓])		Mid-clause pause freq. (Blocked [↑])
Day 3	Mid-clause pause freq. (Blocked [↓])	Repetition (Blocked [↑] , Interleaved [↓])	
Pre-Post			Clause-final pause dur. (Blocked [↓])

Note. Upward arrows indicate an increase in fluency measure corresponding to a higher aptitude score, whereas downward arrows indicate a decrease in fluency measure corresponding to a higher aptitude score.

Repair fluency: Higher PSTM and attention control predict more frequent repairs in the posttest relative to pretest

Although repair frequency (i.e. reformulations and replacements) was not influenced by blocked and interleaved practice schedules (see Suzuki, 2020), repair frequency changes are sensitive to PSTM and attention control. In the current study, learners with higher PSTM and higher attention control made *more* repairs in the posttest relative to pretest.³ Given that the findings yielded by extant descriptive cross-sectional research suggest that higher attention control is associated with *fewer* repairs in speech performance (Zuniga & Simard, 2019), the opposite direction of the current results is intriguing from a developmental perspective.

As delineated above, greater number of repairs can be interpreted as an indicator of increased speech monitoring (Kormos, 1999; Lambert et al., 2020) or heightened attention on accuracy (Brown Nielson & DeKeyser, 2019; Gilabert, 2007). As a case in point, more frequent self-repairs were found to be a good predictor of L2 speaking proficiency gains during a semester study abroad (Golonka, 2006). Because the current L2 learners presumably received extensive grammar instruction in the EFL context for at least 6 years before entering university, they tended to self-initiate repairs for more accurate utterances. For instance, in the speech samples provided by the participants, there were many instances of repairs related to morphological features (e.g. “sunglass” to “sunglasses”, “help” to “helped”, “steal” to “stole”). These simpler repairs, as opposed to more complex discourse-level repairs, are commonly observed among EFL learners exhibiting less efficient or automatic linguistic processing capacity (e.g. Kahng, 2014; Chapter 6 in Kormos, 2006). It is thus tentatively argued that the increase in repair frequency may be a sign of positive L2 developmental change (e.g. more frequent speech monitoring) in the current intervention study (see Hanzawa, 2021, who also showed a positive correlation between repair frequency and L2 learning experience outside the classroom). Greater cognitive resources afforded by higher PSTM and better attention control thus seem to have allowed these L2 learners to direct their attentional resources to monitor their speech in an attempt to improve linguistic accuracy.

Selective roles of memory in fluency development through blocked task repetition

In answer to the second research question, five significant ATI effects revealed in the current study are summarized in Table 3. However, four of the five ATI patterns were driven by the association between aptitude and breakdown fluency changes that were primarily observed in the blocked group. The unique role of aptitude in blocked practice is consistent with the findings yielded by previous research on L2 grammar learning (Suzuki et al., 2020). Nonetheless, it is worth noting that Suzuki et al. (2020) examined grammar learning by distinguishing similar relative clause constructions pertaining to the oral picture description. Hence, the nature of target speaking skills in the current intervention (i.e. utterance fluency in narrative tasks) is very different. Exploring the association between aptitude and different breakdown/repair pausing behaviors allowed us to elucidate cognitive processing underlying L2 speech mechanisms (e.g. proceduralization, speech monitoring, serial vs. parallel processing), which are enhanced differentially by the blocked and interleaved task repetition schedules.

The ATI pattern identified in this study pertained to two significant interactions between PSTM and mid-clause pause frequency in the blocked practice group. Specifically, in this group, higher PSTM led to less mid-clause pause frequency during the training sessions on Day 2 and Day 3.⁴ Because the same story was narrated three times on each day in the blocked practice condition, learners with higher PSTM might have been able to reuse the same or similar linguistic items (e.g. single lexical items or chunks) more efficiently without breakdowns in the third performance relative to the first performance on Days 2 and 3. On the other hand, the role of PSTM in the pretest–posttest change was not evident, possibly because learners had to narrate an unfamiliar story in the posttest and could not reuse or transfer the previously encountered linguistic items. Given the short-term fluency training, even learners with higher PSTM might not have been able to parallel process conceptualization and linguistic formulation (e.g. lexical retrieval and syntactic encoding) in a new story narration (Kormos, 2006; Lambert et al., 2017, 2020; Skehan, 2009).

A more complex pattern emerged for the role of associative memory in fluency development in the blocked practice group. Specifically, on Day 2, learners possessing higher associative memory made *more* mid-clause pauses, while they made fewer clause-final pauses on the posttest relative to pretest. Because the LLAMA_B test, which assesses associative memory, required learners to remember word–picture pairs in 2 min for immediate recall, learners who excelled at this test could have also remembered some words and phrases learned and used in the first performance and tried to use (some of) these in the third performance of the day, which might have overloaded their cognitive capacity, eventually leading to more frequent mid-clause pauses. Perhaps, they might have tried to produce different expressions using a relatively new L2 knowledge. However, owing to their superior memory, these learners arguably developed more control over those linguistic patterns as they progressed through the practice, which could be indicated by the shorter clause-final pauses in the posttest relative to the pretest performance. In other words, learners with more highly developed associative memory were able to link the content (proposition) and linguistic constructions more quickly

(efficiently), which may arguably indicate the shift from serial processing to parallel processing of content and language (Kormos, 2006; Lambert et al., 2020; Skehan, 2009). At this point, any interpretation can be provided for this intriguing pattern in that the same memory component differentially affected different dimensions of breakdown fluency in the blocked task repetition. More detailed analyses on linguistic structures used over the repeated performance (e.g. see de Jong & Perfetti, 2011 for a lexical overlap analysis) may offer more satisfactory explanations not only for this particular finding, but also for the tentative interpretations offered above.

The only cross-over interaction detected in the examined dataset pertained to the association between attention control and repetition change on Day 3, suggesting that attention control influenced fluency change in the opposite directions in the blocked and the interleaved practice group. Specifically, higher attention control was associated with fewer repetitions in the interleaved group, whereas higher attention control was associated with more repetitions in the blocked group. Although this cross-over interaction was significant, a correlation between TMT (attention control) scores and repetition change on Day 3 was weak and nonsignificant in the blocked practice group ($r = .23, p = .30$) and was slightly higher, but still weak, in the interleaved practice group ($r = -.33, p = .12$). Although this intriguing ATI pattern needs to be interpreted with caution, it can be argued that greater attentional control may be beneficial in facilitating an aspect of fluency development, particularly in interleaved task repetition. In the interleaved practice group, three different prompts were used for narrative practice. Practicing speaking skills through alternating narrative stories may require shifting one's attention to different aspects of prompts presented on the same day, which might have taxed this attention control ability. As this interpretation remains speculative, this cross-over ATI effect needs to be further attested in future research.

Limitations of the study and future directions

As this study is subject to several limitations, these should be addressed in future research in this domain. First, post hoc power analysis implemented through G*Power (Faul et al., 2007) indicated that the present multiple regression analysis was underpowered ($1-\beta = .58$) given the current sample size to achieve a medium effect size for the predictors in the model. A large sample size is required to achieve high power (e.g. 80 participants for the statistical power of .80) for a potential replication research design.

Second, only three memory components were assessed in this study. Attention control is only one aspect of the multifaceted construct of working memory, and its different aspects such as updating and inhibition control (Linck et al., 2013) should also be examined to explore their role in fluency development. In addition, the NWRT is a measure that tapped not only PSTM but also L1 articulatory skills. There could have been individual differences reflecting articulation skills that were responsible for (some of) the effects of PSTM on fluency development.

Third, because L2 narrative tasks employed for the training and pretest–posttest measures have many other characteristics that demand beyond L2 speech fluency. It is possible that the current aptitude scores were significantly correlated to some aspects of L2 fluency due to the general non-L2 cognitive abilities and L1 speech

styles. In future research, researchers should examine how well the cognitive aptitude measures predict L2 fluency after controlling for individual differences in L1 speech fluency (see Segalowitz, 2010 for further discussion).

Fourth, repair frequency may be further analyzed to capture different types of repair behaviors such as conceptualizer repairs (pragmatic, semantic, and lexical changes) and formulator repairs (grammatical and articulatory encoding) (Zuniga & Simard, 2019). As the current analyses revealed that the examined memory components played different roles in different dimensions of fluency, authors of future research in this field should examine a wider variety of aptitudes and useful fluency measures (e.g. see Segalowitz et al., 2017 for a useful selection procedure of fluency measures).

Lastly, the intervention adopted in the present investigation involved only nine training sessions distributed over a 3-day period. A more longitudinal research design with an intensive training program (e.g. over 20 sessions in 1 month) is thus needed in order to capture the developmental stages of fluency and putative underlying proceduralization and possibly further automatization.

Conclusions

The aim of the current study was to investigate the role of three cognitive aptitude components in fluency development through blocked and interleaved task repetition. Broadly, the current findings contribute to the growing body of L2 research on cognitive aptitude (Wen et al., 2019). A unique aspect of findings is that not all aspects of fluency are susceptible to cognitive aptitude components. Specifically, no systematic relationship was found between the aptitude components examined in this study and speed fluency (articulation rate) change. However, PSTM and associative memory predicted different aspects of breakdown fluency changes. While PSTM played a significant role in mid-clause pause decrease during the training phase, associative memory was significantly associated with the increase in clause-final pauses in the posttest relative to pretest. This finding suggests that the location of pauses may distinguish different cognitive processes (e.g. linguistic encoding vs. conceptualization such as microplanning) enhanced by different memory components. Although the role of attention control was limited (no effect was observed during the training), attention control, as well as PSTM, was implicated in increasing learners' attention to speech monitoring indicated by greater repair frequency.

Furthermore, several intriguing ATI patterns were identified, suggesting that individual aptitudes play a selective role in blocked task repetition. Most notably, while PSTM contributed to the decrease in mid-clause pause frequency during Day 2 and Day 3 training sessions in the blocked task repetition condition, associative memory contributed to shorter clause-final pause duration in the posttest relative to the pretest. Tentative pedagogical implications from the current ATI patterns may be drawn to optimize L2 speaking practice (DeKeyser, 2019; Vatz et al., 2013). Because the effects of interleaved practice on fluency development are not influenced by individuals' aptitudes at least examined in the study, interleaved practice may be suitable for L2 learners with lower aptitudes. In other words, interleaved

practice may neutralize the levels of aptitudes (Suzuki et al., 2020). In contrast, learners with higher aptitudes (PSTM and associative memory) may take advantage of their strengths and benefit from engaging in blocked task repetition practice most. This is one of the first attempts to investigate the predictive role of aptitude in fluency training intervention research and the findings yielded have generated several interesting questions regarding potential ATI patterns that should be addressed in future studies.

Supplementary material. To view supplementary material for this article, please visit <https://doi.org/10.1017/S0142716421000187>.

Acknowledgments. This study was supported by Grant-in-Aid for Scientific Research (KAKENHI) from the Japan Society for the Promotion of Science (JP18K12470). I would like to express my gratitude to Nel De Jong, Shungo Suzuki, and Keiko Hanzawa for sharing their insight and expertise that greatly assisted this research. I am grateful to Maki Shimizu, Akihiro Yamaguchi, and Chie Takeda for allowing me to use their tests. I also thank Atsushi Miura, Misaki Kuratsubo, Miyu Koyama, Taeko Hosaka, Kazuma Arai for their dedicated assistance in data collection and coding.

Notes

1 The two groups were comparable in terms of three memory scores, according to the independent samples *t*-tests ($p > .10$).

2 Higher associative memory (LLAMA_B) was related to *lower* clause-final pause frequency in the earlier training phase (i.e. Day 1). As it is presently uncertain why learners that scored higher on associative memory tests initially tended to make fewer clause-final pauses, but their number subsequently increased, this phenomenon should be explored in future research.

3 Note that LLAMA_B score (as a measure of associative memory) was a marginally significant predictor of a decrease in repair frequency ($B = -0.11$, $p = .06$), whereas the other two aptitude components exerted positive influence. These findings warrant further investigation.

4 On Day 1, the interaction between NWRT (PSTM) and Group also exerted marginally significant influence on the mid-clause pause duration change ($B = 0.01$, $p = .098$).

References

- Abe, M., Suzuki, K., & Okada, K. (2004). Normative data on tests for frontal lobe functions: Trail making test, verbal fluency, wisconsin card sorting test (keio version). [前頭葉機能検査における中高年健常日本人データの検討: Trail making Test, 語列挙, ウィスコンシンカード分類検査(慶応版)]. *Brain and Nerve*, *56*, 567–574.
- Ahmadian, M. J., & Tavakoli, M. (2011). The effects of simultaneous use of careful online planning and task repetition on accuracy, complexity, and fluency in efl learners' oral production. *Language Teaching Research*, *15*, 35–59. doi: [10.1177/1362168810383329](https://doi.org/10.1177/1362168810383329)
- Baddeley, A. D. (2012). Working memory: Theories, models, and controversies. *Annual review of psychology*, *63*, 1–29. doi: [10.1146/annurev-psych-120710-100422](https://doi.org/10.1146/annurev-psych-120710-100422)
- Brown Nielson, K., & DeKeyser, R. (2019). Working memory and planning time as predictors of fluency and accuracy. *Journal of Second Language Studies*, *2*, 281–316. doi: [10.1075/jsls.19004.bro](https://doi.org/10.1075/jsls.19004.bro)
- Bygate, M. (1996). Effects of task repetition: Appraising the developing language of learners. In J. Willis & D. Willis (Eds.), *Challenge and change in language teaching* (pp. 136–146). Heinemann.
- Bygate, M. (2001). Effects of task repetition on the structure and control of oral language. In P. S. M. Bygate, & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching and testing* (pp. 23–48). Pearson Longman.
- Bygate, M. (2018). *Learning language through task repetition*. Amsterdam, the Netherlands: John Benjamins Publishing Company.

- Cronbach, L. J., & Snow, R. E.** (1977). *Aptitudes and instructional methods: A handbook for research on interactions*. Irvington Pub.
- Dąbrowska, E.** (2019). Experience, aptitude, and individual differences in linguistic attainment: A comparison of native and nonnative speakers. *Language Learning*, **69**, 72–100. doi: [10.1111/lang.12323](https://doi.org/10.1111/lang.12323)
- de Jong, N., & Perfetti, C. A.** (2011). Fluency training in the ESL classroom: An experimental study of fluency development and proceduralization. *Language Learning*, **61**, 533–568. doi: [10.1111/j.1467-9922.2010.00620.x](https://doi.org/10.1111/j.1467-9922.2010.00620.x)
- de Jong, N., & Tillman, P. C.** (2018). Grammatical structures and oral fluency in immediate task repetition: Trigrams across repeated performances. In M. Bygate (Ed.), *Language learning through task repetition* (pp. 43–73). Amsterdam, The Netherlands: 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**, 387–404. doi: [10.1177/1362168815606161](https://doi.org/10.1177/1362168815606161)
- 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**, 113–132. doi: [10.1515/iral-2016-9993](https://doi.org/10.1515/iral-2016-9993)
- de Jong, N. H., & Mora, J. C.** (2019). Does having good articulatory skills lead to more fluent speech in first and second languages? *Studies in Second Language Acquisition*, **41**, 227–239. doi: [10.1017/S0272263117000389](https://doi.org/10.1017/S0272263117000389)
- de Jong, N. H., Steinel, M. P., Florijn, A., Schoonen, R., & Hulstijn, J.** (2013). Linguistic skills and speaking fluency in a second language. *Applied Psycholinguistics*, **34**, 893–916. doi: [10.1017/S0142716412000069](https://doi.org/10.1017/S0142716412000069)
- DeKeyser, R. M.** (2012). Interactions between individual differences, treatments, and structures in sla. *Language Learning*, **62**, 189–200. doi: [10.1111/j.1467-9922.2012.00712.x](https://doi.org/10.1111/j.1467-9922.2012.00712.x)
- DeKeyser, R. M.** (2019). Introduction to the special issue of aptitude-treatment interaction in second language learning. *Journal of Second Language Studies*, **2**, 165–168. doi: [10.1075/jsls.00007.int](https://doi.org/10.1075/jsls.00007.int)
- Engle, R. W.** (2002). Working memory capacity as executive attention. *Current directions in psychological science*, **11**, 19–23. doi: [10.1111/1467-8721.00160](https://doi.org/10.1111/1467-8721.00160)
- Erlam, R.** (2005). Language aptitude and its relationship to instructional effectiveness in second language acquisition. *Language Teaching Research*, **9**, 147–172. doi: [10.1191/1362168805lr161oa](https://doi.org/10.1191/1362168805lr161oa)
- 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.
- Field, A.** (2009). *Discovering statistics using spss*: Sage Publications Limited.
- Foster, P., Tonkyn, A., & Wigglesworth, G.** (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics*, **21**, 354–375. doi: [10.1093/applin/21.3.354](https://doi.org/10.1093/applin/21.3.354)
- Fukuta, J.** (2016). Effects of task repetition on learners' attention orientation in L2 oral production. *Language Teaching Research*, **20**, 321–340. doi: [10.1177/1362168815570142](https://doi.org/10.1177/1362168815570142)
- Gilbert, R.** (2007). Effects of manipulating task complexity on self-repairs during L2 oral production. *International Review of Applied Linguistics in Language Teaching*, **45**, 215–240. doi: [10.1515/iral.2007.010](https://doi.org/10.1515/iral.2007.010)
- Golonka, E. M.** (2006). Predictors revised: Linguistic knowledge and metalinguistic awareness in second language gain in russian. *The Modern Language Journal*, **90**, 496–505.
- Goto, K., Maki, H., & Kasai, C.** (2010). The minimal English test: A new method to measure English as a second language proficiency. *Evaluation & Research in Education*, **23**, 91–104. doi: [10.1080/09500791003734670](https://doi.org/10.1080/09500791003734670)
- Granena, G.** (2013). Individual differences in sequence learning ability and second language acquisition in early childhood and adulthood. *Language Learning*, **63**, 665–703. doi: [10.1111/lang.12018](https://doi.org/10.1111/lang.12018)
- Granena, G., & Yilmaz, Y.** (2019). Phonological short-term memory capacity and L2 oral performance. *Journal of Second Language Studies*, **2**, 317–335. doi: [10.1075/jsls.19005.gra](https://doi.org/10.1075/jsls.19005.gra)
- Grey, S., Cox, J. G., Serafini, E. J., & Sanz, C.** (2015). The role of individual differences in the study abroad context: Cognitive capacity and language development during short-term intensive language exposure. *The Modern Language Journal*, **99**, 137–157. doi: [10.1111/modl.12190](https://doi.org/10.1111/modl.12190)
- Hanzawa, K.** (2021). Development of second language speech fluency in foreign language classrooms: A longitudinal study. *Language Teaching Research, Early View*, 13621688211008693. doi: [10.1177/13621688211008693](https://doi.org/10.1177/13621688211008693)
- Heaton, J. B.** (1996). *Composition through pictures*. Longman.

- Housen, A., & Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics*, *30*, 461–473. doi: [10.1093/applin/amp048](https://doi.org/10.1093/applin/amp048)
- Huang, T., Loerts, H., & Steinkrauss, R. (2020). The impact of second- and third-language learning on language aptitude and working memory. *International Journal of Bilingual Education and Bilingualism*, *1*–17. doi: [10.1080/13670050.2019.1703894](https://doi.org/10.1080/13670050.2019.1703894)
- Hwu, F., Pan, W., & Sun, S. (2014). Aptitude-treatment interaction effects on explicit rule learning: A latent growth curve analysis. *Language Teaching Research*, *18*(3), 294–319. doi: [10.1177/1362168813510381](https://doi.org/10.1177/1362168813510381)
- Kahng, J. (2014). Exploring utterance and cognitive fluency of L1 and L2 English speakers: Temporal measures and stimulated recall. *Language Learning*, *64*, 809–854. doi: [10.1111/lang.12084](https://doi.org/10.1111/lang.12084)
- Kahng, J. (2018). The effect of pause location on perceived fluency. *Applied Psycholinguistics*, *39*, 569–591. doi: [10.1017/S0142716417000534](https://doi.org/10.1017/S0142716417000534)
- Keith, T. Z. (2015). *Multiple regression and beyond: An introduction to multiple regression and structural equation modeling*. Routledge.
- Koponen, M., & Riggenbach, H. (2000). Overview: Varying perspectives on fluency. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 5–24). Michigan: University of Michigan Press.
- Kormos, J. (1999). Monitoring and self-repair in L2. *Language Learning*, *49*, 303–342. doi: [10.1111/0023-8333.00090](https://doi.org/10.1111/0023-8333.00090)
- Kormos, J. (2006). *Speech production and second language acquisition*. Routledge.
- Lambert, C., Aubrey, S., & Leeming, P. (2020). Task preparation and second language speech production. *TESOL Quarterly, Early View*. doi: [10.1002/tesq.598](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*, 167–196. doi: [10.1017/S0272263116000085](https://doi.org/10.1017/S0272263116000085)
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. MIT Press.
- Li, S. (2013). The interactions between the effects of implicit and explicit feedback and individual differences in language analytic ability and working memory. *The Modern Language Journal*, *97*(3), 634–654. doi: [10.1111/j.1540-4781.2013.12030.x](https://doi.org/10.1111/j.1540-4781.2013.12030.x)
- Li, S. (2016). The construct validity of language aptitude: A meta-analysis. *Studies in Second Language Acquisition*, *38*, 801–842.
- Linck, J. A., Hughes, M. M., Campbell, S. G., Silbert, N. H., Tare, M., Jackson, S. R., . . . Doughty, C. J. (2013). Hi-lab: A new measure of aptitude for high-level language proficiency. *Language Learning*, *63*, 530–566. doi: [10.1111/lang.12011](https://doi.org/10.1111/lang.12011)
- Lynch, T., & Maclean, J. (2000). Exploring the benefits of task repetition and recycling for classroom language learning. *Language Teaching Research*, *4*, 221–250. doi: [10.1177%2F136216880000400303](https://doi.org/10.1177%2F136216880000400303)
- Marsden, E., Mackey, A., & Plonsky, L. (2016). The iris repository: Advancing research practice and methodology. In A. Mackey & E. Marsden (Eds.), *Advancing methodology and practice: The iris repository of instruments for research into second languages* (pp. 1–21). Routledge.
- Meara, P. M. (2005). *Llama language aptitude tests: The manual*. Lognostics. Retrieved from http://www.lognostics.co.uk/tools/llama/llama_manual.pdf
- Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions four general conclusions. *Current Directions in Psychological Science*, *21*, 8–14.
- Muñoz, C. (2014). The association between aptitude components and language skills in young learners. In M. Pawlak & L. Aronin (Eds.), *Essential topics in applied linguistics and multilingualism: Studies in honor of david singleton* (pp. 51–68). Springer International Publishing.
- O'Brien, I., Segalowitz, N., Freed, B. F., & Collentine, J. (2007). Phonological memory predicts second language oral fluency gains in adults. *Studies in Second Language Acquisition*, *29*, 557–581.
- Ranta, L. (2002). The role of learners' language analytic ability in the communicative classroom. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 159–179). John Benjamins.
- Reitan, R. M., & Wolfson, D. (1985). *The halstead-reitan neuropsychological test battery: Theory and clinical interpretation* (Vol. 4): Reitan Neuropsychology.
- Robinson, P. (1997). Individual differences and the fundamental similarity of implicit and explicit adult sla. *Language Learning*, *47*, 45–99. doi: [10.1111/0023-8333.21997002](https://doi.org/10.1111/0023-8333.21997002)
- Roger, G., & Carmen, M. (2010). Differences in attainment and performance in a foreign language: The role of working memory capacity. *International Journal of English Studies*, *10*. doi: [10.6018/ijes/2010/1/113961](https://doi.org/10.6018/ijes/2010/1/113961)

- Saito, K. (2017). Effects of sound, vocabulary, and grammar learning aptitude on adult second language speech attainment in foreign language classrooms. *Language Learning*, *67*, 665–693. doi: [10.1111/lang.12244](https://doi.org/10.1111/lang.12244)
- Saito, K. (2019). The role of aptitude in second language segmental learning: The case of Japanese learners' English /ɹ/ pronunciation attainment in classroom settings. *Applied Psycholinguistics*, *40*, 183–204. doi: [10.1017/S0142716418000528](https://doi.org/10.1017/S0142716418000528)
- Sana, F., Yan, V. X., & Kim, J. A. (2017). Study sequence matters for the inductive learning of cognitive concepts. *Journal of Educational Psychology*, *109*, 84–98. doi: [10.1037/edu0000119](https://doi.org/10.1037/edu0000119)
- Segalowitz, N., French, L., & Guay, J.-D. (2017). What features best characterize adult second language utterance fluency and what do they reveal about fluency gains in short-term immersion? *Canadian Journal of Applied Linguistics*, *20*, 90–116. doi: [10.7202/1050813ar](https://doi.org/10.7202/1050813ar)
- Segalowitz, N. S. (2010). *Cognitive bases of second language fluency*. Taylor & Francis.
- Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied linguistics*, *30*, 510–532. doi: [10.1093/applin/amp047](https://doi.org/10.1093/applin/amp047)
- Skehan, P. (2016). Foreign language aptitude, acquisitional sequences, and psycholinguistic processes. In G. Granena, D. O. Jackson, & Y. Yilmaz (Eds.), *Cognitive individual differences in second language processing and acquisition* (pp. 17–40). John Benjamins.
- Skehan, P., Foster, P., & Shum, S. (2016). Ladders and snakes in second language fluency. *International Review of Applied Linguistics in Language Teaching*, *54*, 97–111. doi: [10.1515/iral-2016-9992](https://doi.org/10.1515/iral-2016-9992)
- Soper, D. S. (2006). Interaction! Retrieved from <https://www.danielsooper.com/Interaction/default.aspx>
- Suzuki, Y. (2019). Individualization of practice distribution in second language grammar learning: A role of metalinguistic rule rehearsal ability and working memory capacity. *Journal of Second Language Studies*, *2*, 170–197. doi: [10.1075/jsls.18023.suz](https://doi.org/10.1075/jsls.18023.suz)
- Suzuki, Y. (2020). Optimizing fluency training for speaking skills transfer: Comparing the effects of blocked and interleaved task repetition. *Language Learning, Early View*. doi: [10.1111/lang.12433](https://doi.org/10.1111/lang.12433)
- Suzuki, Y., & DeKeyser, R. M. (2017a). Exploratory research on L2 practice distribution: An aptitude × treatment interaction. *Applied Psycholinguistics*, *38*, 27–56. doi: [10.1017/S0142716416000084](https://doi.org/10.1017/S0142716416000084)
- Suzuki, Y., & DeKeyser, R. M. (2017b). The interface of explicit and implicit knowledge in a second language: Insights from individual differences in cognitive aptitudes. *Language Learning*, *67*, 747–790. doi: [10.1111/lang.12241](https://doi.org/10.1111/lang.12241)
- Suzuki, Y., Yokosawa, S., & Aline, D. (2020). The role of working memory in blocked and interleaved grammar practice: Proceduralization of L2 syntax. *Language Teaching Research, Early View*. doi: [10.1177/1362168820913985](https://doi.org/10.1177/1362168820913985)
- Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Boston: Pearson Education.
- Tare, M., Golonka, E., Lancaster, A. K., Bonilla, C., Doughty, C. J., Belnap, R. K., & Jackson, S. R. (2018). The role of cognitive aptitudes in a study abroad language-learning environment. In C. Sanz & A. Morales-Front (Eds.), *The routledge handbook of study abroad research and practice* (pp. 406–420). Routledge.
- Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure and performance testing. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 239–273). John Benjamins.
- Thai, C., & Boers, F. (2016). Repeating a monologue under increasing time pressure: Effects on fluency, complexity, and accuracy. *TESOL Quarterly*, *50*, 369–393. doi: [10.1002/tesq.232](https://doi.org/10.1002/tesq.232)
- Trofimovich, P., Ammar, A., & Gatbonton, E. (2007). How effective are recasts? The role of attention, memory, and analytical ability. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A series of empirical studies* (pp. 171–195). Oxford University Press.
- Vatz, K., Tare, M., Jackson, S., & Doughty, C. (2013). Aptitude-treatment interaction studies in second language acquisition: Findings and methodology. In G. Granena & M. H. Long (Eds.), *Sensitive periods, language aptitude, and ultimate L2 attainment* (pp. 273–292). John Benjamin's Publishing.
- Wen, Z. (2016). Phonological and executive working memory in L2 task-based speech planning and performance. *The Language Learning Journal*, *44*, 418–435. doi: [10.1080/09571736.2016.1227220](https://doi.org/10.1080/09571736.2016.1227220)
- Wen, Z., Biedroń, A., & Skehan, P. (2017). Foreign language aptitude theory: Yesterday, today and tomorrow. *Language Teaching*, *50*, 1–31. doi: [10.1017/S0261444816000276](https://doi.org/10.1017/S0261444816000276)
- Wen, Z., Skehan, P., Biedroń, A., Li, S., & Sparks, R. L. (2019). *Language aptitude: Advancing theory, testing, research and practice*. Routledge.

- Yamaguchi, A., & Shimizu, M.** (2011). The relationship between phonological working memory and English proficiency of Japanese university efl learners (2). [音韻的作動記憶と英語熟達度の関係の検討(2)日本人大学生を対象として]. *Annual report of the Faculty of Education, Gunma University. Cultural Science Series*, **60**, 233–242.
- Yilmaz, Y.** (2013). Relative effects of explicit and implicit feedback: The role of working memory capacity and language analytic ability. *Applied Linguistics*, **34**, 344–368. doi: [10.1093/applin/ams044](https://doi.org/10.1093/applin/ams044)
- Zuniga, M., & Simard, D.** (2019). Factors influencing L2 self-repair behavior: The role of L2 proficiency, attentional control and L1 self-repair behavior. *Journal of Psycholinguistic Research*, **48**, 43–59. doi: [10.1007/s10936-018-9587-2](https://doi.org/10.1007/s10936-018-9587-2)

Cite this article: Suzuki, Y. (2021). 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**, 969–995. <https://doi.org/10.1017/S0142716421000187>