Does automaticity in lexical and grammatical processing predict utterance fluency development?

A six-month longitudinal study in Japanese EFL context

Keiko Hanzawa and Yuichi Suzuki
Tokyo University of Science, Japan | Kanagawa University, Japan

In order to understand the development of automatization in second language (L2) acquisition and its role in speaking fluency development, 39 English as foreign language (EFL) learners performed a speaking task (subsequently analyzed for utterance fluency) as well as semantic classification (lexical processing measure) and maze (grammar processing measure) tasks at two time points (Time 1 and Time 2) over six months. The findings indicate that learners significantly speeded up their lexical and grammar processing, as reflected in faster reaction time (RT). However, only lexical processing showed a propensity for more stable processing at Time 1, as indicated by coefficient of variance (CV). Furthermore, multiple regression analyses revealed that more automatized grammar processing (faster RT and smaller CV) at Time 1 significantly predicted a larger reduction in mid-clause pause duration and frequency, respectively. These findings underscore the importance of automatization in grammar processing for developing fluency in an EFL context.

Keywords: cognitive fluency, utterance fluency, automatization, lexical processing, grammar processing, longitudinal study

1. Introduction

Speaking a second language (L2) fluently is a major goal for many L2 learners. Because producing fluent speech requires a myriad of components to be executed and coordinated efficiently, it is an impressive undertaking to achieve L2 fluency particularly in a foreign language context, where L2 exposure is limited relative to

https://doi.org/10.1075/jsls.22007.han
Journal of Second Language Studies 6:2 (2023), pp. 292–316. ISSN 2542-3835 | E-ISSN 2542-3843

second language context (Collins & Muñoz, 2016). Therefore, obtaining a better understanding of how L2 fluency is achieved in the foreign language context is of importance for L2 researchers and practitioners.

Although in non-technical sense fluency can simply refer to speaking smoothly without serious disruptions, fluency in L2 research is considered a multifaceted construct and is researched more comprehensively. According to the seminal work by Segalowitz (2010, 2016), there are three dimensions in L2 fluency: (a) perceived fluency, (b) utterance fluency, and (c) cognitive fluency. Perceived fluency is a subjective impression typically elicited by judgement about speakers' speech. Utterance fluency refers to "the set of objectively determined timing, pausing/hesitation, and repair features of the utterance" (Segalowitz, 2010, p. 49), which is supported by cognitive fluency. Cognitive fluency concerns "the speaker's ability to efficiently mobilize and integrate the underlying cognitive processes responsible for producing utterances ~~with the characteristics that they have~~" (p. 49). The current study focuses on the latter two dimensions of fluency: utterance fluency and cognitive fluency.

We situate the current study in the framework of broader skill-acquisition literature (e.g., DeKeyser, 2007, 2017 w th the presupposition that cognitive fluency supports utterance fluency (e.g., Segalowitz & Freed, 2004). In this context, cognitive fluency refers to L2-specific processing skills and the efficiency for which lexical and grammatical resources can be deployed (e.g., de Jong, Steinel, Florijn, Schoonen, & Hulstijn, 2013; Segalowitz, 2010, 2016). This processing skill view of cognitive fluency can be tied to the concept of automaticity (e.g., DeKeyser, 2001; Segalowitz, 2003). Based on L1 and L2 speech models, three major stages are postulated in the L2 speech processing: conceptualization, formulation, and articulation (Kormos, 2006; Levelt, 1989). Automatization is presumably most relevant for the formulation stage, as a part of which the preverbal plans generated in the conceptualization stage are linguistically encoded using lexical and syntactic resources stored in speakers' mental lexicon. Hence, investigating automatization in lexical and syntactic processing is important to advance our understanding of Recently 2 speech is coordinated efficiently and fluently.

~~Relatedly~~, researchers have started to elucidate the complexities of L2 speech mechanisms by examining the link between cognitive fluency (including automaticity in lexical and grammatical processing) and utterance fluency using a cross-sectional research design (de Jong et al., illuminate 020; S. Suzuki & Kormos, 2022). Other researchers, aiming to ~~elucidate~~ L2 speaking skill development, have investigated the role of automatization in utterance fluency development in study abroad contexts (i.e., an acquisition-rich environment for L2 speaking) (Leonard & Shea, 2017; Segalowitz & Freed, 2004). However, the developmental trajectory of automatization and its linkage to utterance fluency are underexplored in foreign language contexts where L2 exposure and opportunities to use L2 are more limited.

In order to better understand automatization (i.e., cognitive fluency) and L2 utterance development in foreign language contexts, in this study, we focused on the development of automatization in lexical and grammatical knowledge (i.e., aspects of cognitive fluency) and its role in utterance fluency development among English learners in Japan over six months. Our findings revealed that, when changes in cognitive and utterance fluency were examined longitudinally, automatization of lexical and grammatical knowledge followed different learning trajectories. Furthermore, the extent to which grammatical knowledge was automatized at the onset of the experiment significantly predicted a critical aspect of utterance fluency development.

2. Literature review

2.1 Automatization in L2 learning: Fast and stable execution of L2 skills

Cognitive psychology research posits that automaticity is a hallmark of well-learned knowledge and skills (e.g., Anderson et al., 2004; Shiffrin & Schneider, 1977). In L2 research, automaticity is typically defined as a fast, stable, ballistic, effortless, and unconscious process (Segalowitz, 2003). This definition suggests that automaticity consists of multiple components (Moors, 2016), each of which can be evaluated separately and play differential roles in fluent speech processes. In the current study, we specifically focus on the two commonly researched dimensions of automatic processing: (a) speed and (b) stability.

L2 researchers have proposed that fast and stable L2 processing can differ in significant ways. According to Segalowitz and Segalowitz (1993), executing a L2 skill faster reflects a *quantitative* change, whereby a pre-established process is simply accelerated. In contrast, an L2 skill becoming more stable entails a *qualitative* shift, which involves restructuring (some of) the pre-established processes. For instance, in lexical access, novice learners largely use L1 translation to understand semantic components of L2 speech. As they are gaining skills in L2 but are still in the beginning stage, the L1 translation process can merely become faster (i.e., L2 processing speeds up). However, in later stages, the ~~understanding~~ process is restructured by eliminating the L1 translation process, allowing the meaning of L2 words to be directly accessed (indicating stability of L2 processing). In a similar vein, while faster grammar processing may involve speeding up of L1-based sentence processing and/or use of declarative knowledge (e.g., metalinguistic knowledge of rules), more stable grammar processing may involve the elimination of those processes and the establishment of a more efficient process. Segalowitz and Segalowitz (1993) claim that automaticity in a technical sense uniquely refers to the latter, as opposed to mere speed-up processing.

Processing speed and stability in L2 skills can be measured utilizing reaction-time (RT) psycholinguistic tasks. RT is a gold standard measure of processing speed because it reflects how quickly a given process is executed. Stable processing (a potential index of restructuring) is, in contrast, assessed by the coefficient of variance (CV), which is computed by dividing the mean standard deviation (SD) of RT for each individual by his/her mean RT (Segalowitz & Segalowitz, 1993). These two measures have been utilized to study the extent to which L2 learners automatize their L2 lexical and grammatical skills. For evaluating the degree of automaticity, three criteria have been proposed (e.g., Hulstijn, Van Gelderen, & Schoonen, 2009; Lim & Godfroid, 2015; Y. Suzuki & Sunada, 2018): (1) faster RT, (2) smaller CV, and (3) a positive RT–CV correlation. The third criterion is particularly important for determining if restructuring of L2 processing has occurred (see Segalowitz & Segalowitz, 1993 for details). When L2 skill is practiced, both mean RT and SD usually decrease proportionally. In this case, CV remains unchanged and no positive correlation between RT and CV is expected. However, extensive L2 skills practice can potentially lead to a disproportional reduction in SD relative to RT (e.g., possibly due to an elimination of redundant processing components such as L1 translation and access to metalinguistic knowledge), which results in smaller CV and yields a positive correlation between RT and CV. Hence, the criterion of positive RT–CV correlation can be considered as a "litmus test" for automatization in the sense of restructuring (Hulstijn et al., 2009).

Both RT and CV have been adopted to assess the level of L2 abilities pertaining to lexical processing (e.g., Akamatsu, 2008; Elgort, 2011; Hui & Godfroid, 2020; Phillips, Segalowitz, O'Brien, & Yamasaki, 2004; Segalowitz & Segalowitz, 1993; see Hulstijin et al., 2009 for a detailed review) and grammar processing (Lim & Godfroid, 2015; McManus & Marsden, 2018; Pili-Moss, Brill-Schuetz, Faretta-Stutenberg, & Morgan-Short, 2019; Y. Suzuki, 2018; Y. Suzuki & Sunada, 2018). For instance, propensities for automatization based on (some of) the criteria above (i.e., faster RT, smaller CV, and positive RT–CV correlation) were found for more proficient learners (e.g., Lim & Godfroid, 2015; Segalowitz & Segalowitz, 1993) as well as learners with experience in L2 immersion contexts (Y. Suzuki & Sunada, 2018). Furthermore, in an instructed second language acquisition research, both RT and CV have been proven useful for measuring the learning gains. For instance, for lexical items and grammatical structures that were practiced deliberately, faster RT and smaller CV, and/or positive relationships between them, have been reported (e.g., Elgort, 2011; McManus & Marsden, 2018; Y. Suzuki, 2018).

Although researchers are increasingly relying on RT and CV in their L2 skill acquisition studies, the utility of CV as a predictor of L2 development has not been extensively studied (see Hulstijn et al., 2009; Lim & Godfroid, 2015; Y. Suzuki & Sunada, 2018 for further discussion). Furthermore, in the previous studies, lexical and grammar processing skills of the same learners have rarely been examined *simultaneously* as a part of a longitudinal research design, even though this would allow for a better understanding of the development trajectory of different L2 skills, as well as enable the utility of CV as an indicator of L2 skill acquisition to be evaluated from the automaticity perspective. Due to the lack of empirical investigations, it remains unknown how lexical and grammar processing skills develop over time. Yet, given the nature of lexical learning involving simpler form-meaning mapping, automatization of lexical processing (e.g., of frequent words) is more achievable than that of grammar processing skills involving interpretations and parsing of abstract syntactic structures.

2.2 The role of automatization in utterance fluency development

Prompted by the Tavakoli and Skehan's (2005) seminal work, utterance fluency is typically analyzed in terms of speed, breakdown, and repair fluency. Following Segalowit's (2010) suggestion that utterance fluency is underlined by cognitive fluency, many L2 researchers have attempted to explore how cognitive fluency (automatization) relates to the established utterance fluency measures.

An emerging line of research, typically employing a cross-sectional research design, has yielded some valuable findings regarding the relationship between different utterance fluency measures and L2-specific knowledge and cognitive processes (de Jong et al., 2013; Kahng, 2020; S. Suzuki & Kormos, 2022). A most comprehensive study to date was conducted by S. Suzuki and Kormos (2022), who investigated the relationship between automaticity of phonological, lexical, and grammatical skills and utterance fluency. These authors recruited 128 English-as-foreign-language learners at a Japanese university and administered a battery of language tests, some of which aimed to measure automatization (e.g., picture naming task, maze task), as well as speaking tasks. The results yielded by structural equation modeling revealed that automaticity of L2 syntactic knowledge (syntactic processing speed) was significantly related to speed fluency and breakdown fluency, but not to repair fluency. Furthermore, based on the strengths of factor loadings, the authors argued that the mid-clause pause frequency is a representative component of breakdown fluency. According to S. Suzuki and Kormos (2022), pausing in the middle of clauses presumably reflects disfluency deriving from L2-specific problems, including lack of linguistic resources and automatization.

Other researchers have taken a longitudinal approach and have examined the development of cognitive fluency (automatization) and its role in utterance fluency development in study abroad contexts. A seminal study by Segalowitz and Freed (2004) was conducted with L2 Spanish learners who studied Spanish in classroom (at home) and in study abroad contexts. Both groups took a battery of cognitive fluency tests (lexical processing task and attention control task) as well as a speaking test (Oral Proficiency Interview) twice before and after spending 13 weeks in their respective contexts. While the authors examined several other factors in their work, we focus on the RT and CV related to the lexical processing task, as these measures are most relevant to the current study's goal. The data on the semantic classification task (measures of lexical processing speed) was analyzed for study-abroad ($n = 15$) and at-home leaners ($n = 14$). Both groups showed significant improvements in lexical processing speed (RT) and stability (CV). Furthermore, RT and CV at Time 1 played a significant role in the score gain on the Oral Proficiency Interview test (based on the interviewer's ratings on general speaking proficiency from novice through advanced). However, none of the fluency measures (i.e., speech rate, hesitations-free speech, mean rune length, filler-free speech) were predicted by RT or CV.

In a more recent study, Leonard and Shea (2017) tracked 39 L2 learners in a Spanish-speaking country for three months. These authors administered both lexical and grammatical processing tests as a measure of automaticity and used them as predictors of fluency gains. Their findings revealed that a measure of grammar processing speed (RT derived from the sentence–picture verification) was the strongest predictor of fluency before commencing the study abroad program. However, grammar processing speed was NOT a significant predictor of fluency change during the three-month sojourn. These seemingly contradictory findings may be reconciled by paying attention to the learning context (i.e., study abroad) where opportunities to speak L2 frequently are greater than in at-home or classroom settings. According to Leonard and Shea (2017), this study abroad environment might have helped their learners to improve fluency regardless of their prior grammar processing speed. This interpretation seems plausible, and calls for a further exploration of the potential role of grammar processing for predicting fluency changes in classroom or foreign language settings where fewer opportunities to speak L2 are available. To our knowledge, there is no longitudinal study to date that has investigated the relationship between cognitive and utterance fluency in a foreign language setting.

3. The current study

The aim of the current study is to gain in-depth insight into the development of automaticity (i.e., cognitive fluency) in classroom or foreign language settings and its role in utterance fluency development by going beyond simply describing fluency development (Segalowitz, 2016). This work is a part of a large project that tracks L2 fluency development among Japanese English as foreign language (EFL) learners (see Hanzawa, 2021 for the examination of utterance fluency development). Although Hanzawa (2021) found that utterance fluency developed over the course of one academic year, she did not address how the development of utterance fluency was related to automatization. Therefore, the current study focuses on automatization (a part of cognitive fluency) and the link between automatization and utterance fluency. While Hanzawa (2021) investigated fluency development among Japanese university freshmen at three time points (i.e., at the beginning of the first semester, the end of the first semester, and the end of the second semester), the latter two time points are in focus of the current study, i.e., the end of the first semester (labeled as Time 1; T1) and the end of the second semester (labeled as Time 2; T2), as illustrated in Figure 1.[1]

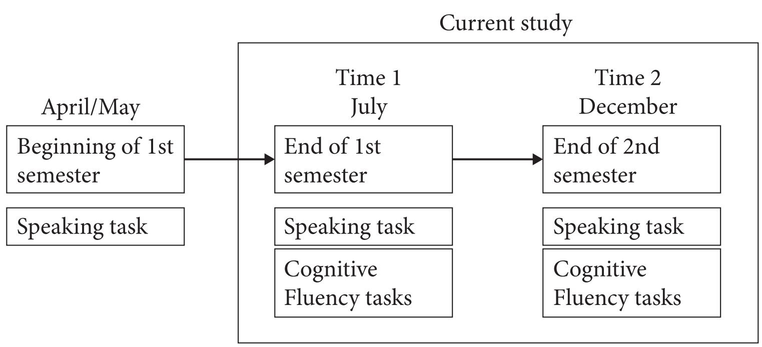

Figure 1. Research design and data collection times
Note. The data for the current study was collected at Time 1 and Time 2.

The following two research questions (RQs) were addressed in the current study:

1. To what extent does automatization happen in lexical and syntactic processing in an EFL setting?

1. Due to technical issues, the cognitive fluency tasks (i.e., semantic classification and maze tasks) were not administered at the beginning of the project in April/May.

1a. Does RT decrease?
1b. Does CV decrease?
1c. Is there a significant correlation between RT and CV at T1 and T2?

2. To what extent does automatization in lexical and syntactic processing (i.e., cognitive fluency) at T1 predict utterance fluency changes over a six-month period?

4. Methods

4.1 Participants

The initial study sample comprised of 50 L2 English learners studying at a Japanese university (mean age = 18.1) who were all first-year students at the onset of the experiment. Participants were native Japanese speakers with both parents being native Japanese speakers. They started learning English in Grade 7 of secondary school and had no prior living abroad experience lasting more than two months. Therefore, they were expected to have similar linguistic backgrounds. At the beginning of the experiment, their general English proficiency was tested by an English proficiency test specifically designed for the university students, indicating that their level fell between B1 (intermediate) and C1 (advanced) on the Common European Framework Reference of Languages (CEFR).

Based on the results of questionnaire administered at T2, the participants took four 1.5-hour English classes per week in the second semester. These classes consisted of three form-focused (English language lessons where input and output in L2 were manipulated to help learners to improve English language skills; Spada, 1997) and one content-based classes (instruction were conducted in English to get learners to acquire academic knowledge and skills such as psychology, math, and history; Lyster & Ballinger, 2011). In addition, the learners engaged in English activities (reading, listening, writing, speaking, vocabulary learning) outside classrooms for 70 hours on average over six months. This suggests that the learners used English in various manners in the EFL setting.

The data from 11 participants for the computerized tasks (maze and semantic classification) was not recorded due to technical problems. Consequently, the analyses reported here pertain to the remaining 39 students.

4.2 Personal narrative task (Utterance Fluency Measure)

A personal narrative task was used to examine the participants' utterance fluency at T1 and T2. In this task, the participants described "the toughest or most challenging event they had experienced in the past few months" in English. They were provided an instruction sheet, which included the prompt question and three guiding questions–"When did it happen?"; "Where did it happen?"; and "Why did you find this experience to be the most challenging?"–which helped them to produce a certain amount of speech without excessive hesitation or disfluency. The participants were provided one minute for planning their narration without taking notes. After the preparation, they were given two minutes to narrate their story while looking at the instruction sheet. The participants were also instructed to begin narrating their story with the prompt sentence ("The most challenging event I have experienced in the past few months…").

The same prompt and procedure were used at T1 and T2 in order to maximize the comparability of test results (e.g., Saito, Suzukida, & Sun, 2018). While repeating the same prompt may lead to practice effect, none of the participants talked about the same story at T1 and T2.

4.3 Cognitive fluency tasks (automaticity measures)

4.3.1 *Semantic classification task*

Lexical processing efficacy was assessed via the semantic classification task, whereby the participants were asked to decide if a word presented on the computer screen refers to a living (e.g., mother) or an inanimate object (e.g., school) as quickly and accurately as possible. Such animacy judgement task has been used to assess learners' semantic processing in previous studies (e.g., Lim & Godfroid, 2015). For the test adopted in the present study, 60 nouns (30 relating to animate and 30 to inanimate objects) were extracted from the maze task. The vocabulary-level coverage was also checked (Cobb, 2002), indicating that 90% of the words were the first 1,000 most frequent words, and the remaining 10% were the second 1,000 most frequent words. The participants were given six practice trials before the actual test to familiarize them with the procedure. They completed the same task at T1 and T2, but the order of items was randomized for every participant at both T1 and T2.

4.3.2 *Maze task*

The maze task assessed sentence processing efficiency. In this task, the participants were required to construct a sentence by selecting options presented on a computer screen as quickly and accurately as possible (see Figure 2). In the first screen, the left option was always the first word of the sentence (e.g., *How* and *x-x-x*). The second screen asked the participants to choose the word from the two options that would correctly follow the previous word by pressing either left (f) or right (j) button (e.g., *enter* and *many*). If the participants responded correctly, the next screen with the next two options appeared, and the process continued until the end of the sentence. On the other hand, if they gave a wrong response, the trial automatically stopped and the next trial started.

Stimuli used in the current study were adapted from the study conducted by Y. Suzuki and Sunada (2018), which included declaratives, wh-questions, relative clauses, and indirect questions (the stimuli were available at: https://www.iris-database.org). Two counterbalanced lists, each of which consisted of 48 sentences, were used at T1 and T2 to minimize the practice effect. According to Y. Suzuki and Sunada (2018), the items in the two lists were found to be of a similar level of difficulty by their participants (i.e., Japanese university students), who have similar linguistic and educational background as those in the current study. After four practice trials, which enabled the participants to familiarize themselves with the task procedure, they completed 48 trials.

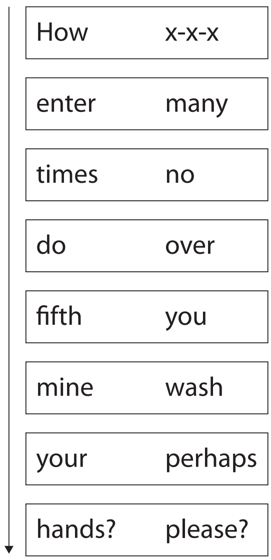

Figure 2. An illustration of the maze task (Answer: "How many times do you wash your hands?")

4.4 Procedure

Data collection was conducted individually at a quiet university's laboratory. The speaking task was administered first using a digital recorder, after which two cognitive fluency tasks were administered through DMDX software (Forster & Forster, 2003), with the semantic classification task preceding the maze task (similar methodological decision was made by Lim & Godfroid, 2015). The same order (i.e., the speaking task followed by the cognitive fluency tests) was adopted at T1 and T2.

4.5 Data coding

4.5.1 *Speaking task*

A total of 78 speech samples (obtained from 39 participants at T1 and T2) were analyzed to examine changes in utterance fluency. The first part of the speech (i.e., the prompt sentence) was removed before fluency coding. The speech was then annotated using a free sound analysis software PRAAT (Boersma & Weenink, 2016). With the aid of the script for automatically detecting pauses (de Jong & Wempe, 2009), the unpruned transcripts were prepared, which were further coded by two trained coders.

To measure utterance fluency, the following six fluency measures were selected ~~extracted~~ to cover speed, breakdown, and repair fluency (Housen & Kuiken, 2009; Skehan, 2009):

Speed fluency

1. Mean syllable duration (MSD) (the total speech time divided by the number of the number of syllables, excluding pauses)

Breakdown fluency

2. Mid-clause pause duration (the mean length of mid-clause filled and unfilled pauses)
3. Clause-final pause duration (the mean length of clause-final filled and unfilled pauses)
4. Mid-clause pause frequency (the number of mid-clause filled and unfilled pauses divided by the total number of syllables)
5. Clause-final pause frequency (the number of clause-final filled and unfilled pauses divided by the total number of syllables)

Repair fluency

6. Repair frequency (the number of self-repairs such as repetitions and self-corrections divided by the total speech duration)

Speed fluency is represented by mean syllable duration (the inverse of articulation rate) (de Jong et al., 2013). For breakdown fluency, four indices were computed: mid-clause and clause-final pause duration and frequency. Pauses were defined as the filled and unfilled (silent) pauses lasting at least 250 ms, excluding micropauses (e.g., bursts before plosive sounds) (Bosker, Pinget, Quene, Sanders,

& de Jong, 2013). They were further coded as mid-clause and clause-final pauses based on pause location using Analysis of Speech (AS) unit (Foster, Tonkyn, & Wigglesworth, 2000). Mid-clause pauses are presumably related to linguistic (e.g., lexical and grammar) processing and clause-final pauses to conceptual processing (Kahng, 2014; Lambert, Aubrey, & Leeming, 2020; Saito, Ilkan, Magne, Tran, & Suzuki, 2018). For repair fluency, all instances of repair frequency (including repetition and self-repairs) were counted (Kormos & Dénes, 2004; S. Suzuki & Kormos, 2020). Note that utterance fluency scores at T1 and T2 are presented in Appendix S1 in the Online Supplementary File (see the original study by Hanzawa, 2021 for further details).

4.5.2 *Semantic classification task*

In the semantic classification task, the responses to each word stimulus were analyzed after excluding outlying responses (indicated by RTs below 300 ms or higher than 3 SD above the group mean of each item). As a result, 1.88% and 1.70% of the responses were excluded from the T1 and T2 dataset, respectively. Mean RT was computed based on the remaining RT data, and CV was obtained by dividing the mean SD of RT for each individual by their mean RT.

4.5.3 *Maze task*

In the maze task, the responses to each word item were collected and analyzed. RT was included in the analysis only when the response was correct. Note that the first item of each sentence was excluded from the analysis, since it was already given to the participants. Outlying responses were identified as RTs below 300 ms or higher than 3 SD above the group mean of each word item. As a result, 1.07% and 1.18% of the responses were excluded from the T1 and T2 dataset, respectively. Individuals' RT and CV were computed in the same way as in the semantic classification task.

4.6 Statistical analysis

To address RQ1a and RQ1b, a series of paired samples t-tests was conducted to examine the development of lexical and grammar processing between T1 and T2. The magnitude of effect size (*d*) was interpreted as small (.40), medium (.70), or large (1.0) according to the L2-specific research benchmark (Plonsky & Oswald, 2014). Furthermore, to answer RQ1c, Pearson's correlation analysis was conducted to examine the correlation between RT and CV for the semantic classification and maze tasks. The effect sizes were interpreted based on the L2-specific research benchmark (Plonsky & Oswald, 2014): small ($r = .25$), medium ($r = .40$), or large ($r = .60$).

Regarding RQ2, hierarchical multiple regression analyses were conducted on the fluency gains (obtained by subtracting the fluency score at T1 from the score obtained at T2, with a greater change signifying more pronounced fluency development) for each of the six fluency measurements. The hierarchical regression analyses consisted of two steps (Keith, 2015). In Step 1, a fluency measure at T1 was entered into the model to control for the baseline performance (Allison, 1990). In Step 2, four cognitive fluency measures (Semantic RT, Semantic CV, Maze RT, and Maze CV) were entered as predictors. The main focus of the current study ~~concerned analyses conducted as~~ was a part of Step 2, because these results pertain to the second research question (predictors of fluency development).

For each regression model, the four assumptions of multiple regression analysis (absence of collinearity, linearity, homoscedasticity, and normality of residuals) were checked. One or two highly influential data points (Cook's $d > 0.25$) were detected and removed from the dataset to which the respective regression models were applied (syllable duration, mid-clause pause frequency and duration, and clause-final pause frequency and duration, see Appendix S2 for the details). After ensuring that all the assumptions were met, multiple regression analyses were conducted.

5. Results

5.1 Development of cognitive fluency (RQ1)

Table 1 and Figure 3 present the paired samples t-test results related to the semantic classification and maze tasks at T1 and T2. For the semantic classification task, the paired samples t-test showed a significant difference in RT between T1 and T2 with a medium effect size, $t(38) = 3.80$, $p < .001$, $d = 0.60$. No significant difference was noted for CV with a negligible effect size, $t(38) = 0.22$, $p = .82$, $d = 0.03$.

In the maze task, the learners showed improvement in the RT at T2 compared with T1. According to the paired samples t-test results, while the RT for the maze task was significantly faster at T2 than T1 with a large effect size, $t(38) = 6.61$, $p < .001$, $d = 1.05$, no significant difference was found in the CV with a negligible effect size, $t(38) = -0.19$, $p = .85$, $d = -0.03$.

Note that the accuracy rates were above 90% for the semantic classification task (T1: $M = 0.97$, $SD = 0.03$; T2: $M = 0.96$, $SD = 0.03$) and the maze task (T1: $M = 0.90$, $SD = 0.06$; T2: $M = 0.92$, $SD = 0.05$). No significant changes occurred for either task: the semantic classification task ($t = 1.19$, $p = .24$, $d = 0.19$) and the maze task ($t = -1.72$, $p = .09$, $d = -0.27$).

Table 1. Paired samples t-test results for the semantic classification and maze tasks

	Time 1		Time 2				
	Mean [95% CI]	SD	Mean [95% CI]	SD	*t*	*p*	*d*
Semantic Classification Task							
RT	703 [679, 727]	73	666 [643, 689]	71	3.80	<.001	0.60
CV	0.23 [0.21, 0.24]	0.05	0.22 [0.20, 0.23]	0.05	0.22	.82	0.03
Maze Task							
RT	1109 [1077, 1142]	100	1007 [973, 1041]	105	6.61	<.001	1.05
CV	0.37 [0.36, 0.38]	0.04	0.38 [0.36, 0.38]	0.04	−0.19	.85	−0.03

Note. CI = confidence interval.

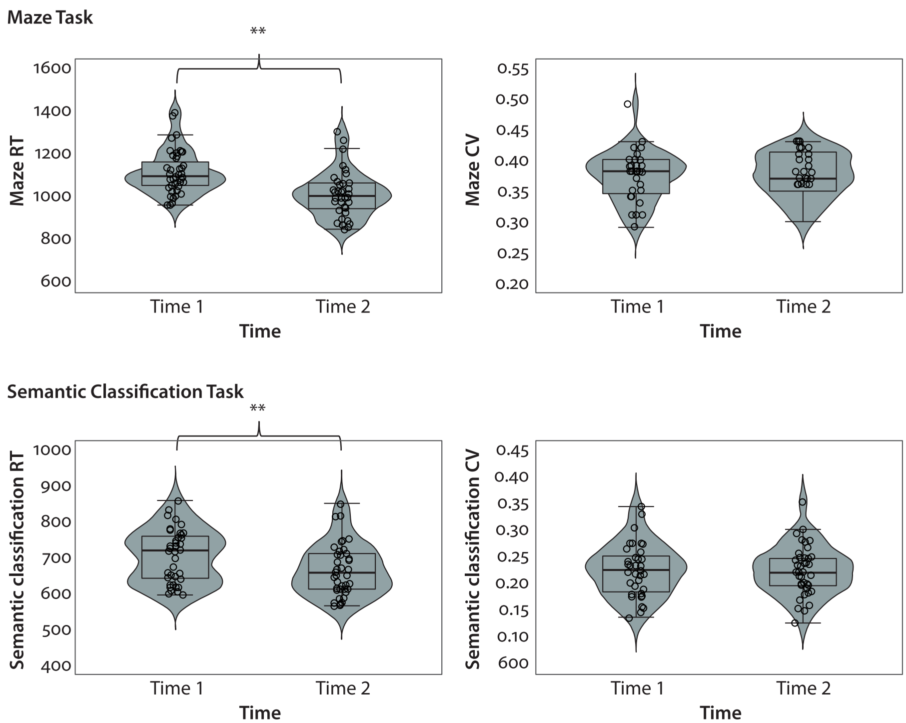

Figure 3. Cognitive fluency scores at Time 1 and Time 2

In order to further evaluate the third automaticity criterion (a positive RT–CV relationship) Pearson's correlation coefficients were computed. As shown in Table 2, the correlation between RT and CV in the semantic classification task at T1 was positive and significant with a medium effect size. Moreover, in the semantic classification task at T2, a marginally significant correlation was found between RT and CV with a small effect size. However, no significant correlation was noted in the maze task at T1 or T2.

Table 2. Correlation between RT and CV for maze and semantic classification tasks at Time 1 and Time 2

	Time 1				Time 2			
			95% CI				95% CI	
	r	*p*	*lower*	*upper*	*r*	*p*	*lower*	*upper*
Semantic Classification Task	.40	.01	.15	.64	.29	.06	.02	.54
Maze Task	−.06	.69	−.31	.19	.07	.63	−.18	.35

5.2 The relationship between cognitive fluency and utterance fluency measures (RQ2)

The hierarchical regression analyses conducted as a part of Step 1 revealed that ~~a~~ fluency ~~measure~~ measures at T1 ~~was~~ were significant in every model, except for mean syllable duration.[2] All coefficients are negative, indicating that less fluent learners at T1 tended to show larger gains in fluency. This finding is consistent with a general observation that learners with lower ability tend to make greater improvements compared to more able leaners (Leonard & Shea, 2017).

The main part of analysis (Step 2) yielded a significant predictor in the regression models related to mid-clause pause frequency and mid-clause pause duration (Table 3). For the mid-clause pause frequency, Maze CV was a significant predictor ($\beta = 0.30$, $p = .03$). For this model, however, the addition of other cognitive fluency scores did not lead to a significant further explanation of variance in mid-clause pause frequency change, $\Delta F(4, 32) = 1.32$, $\Delta R^2 = 10\%$, $p = .28$. This result was most likely due to the redundant predictors. The regression model based solely on Maze CV (i.e., simple linear regression) was significant and accounted for a similar amount of variance in the fluency change, $F(1, 35) = 5.10$, $p = .03$, $\Delta R^2 = 9\%$.

2. The correlation coefficients between fluency at T1 and fluency change scores are presented in see Appendix S3 in the Online Supplementary File.

Similarly, for the mid-clause pause *duration*, Maze RT was a significant predictor ($\beta = 0.31$, $p = .03$). Once again, however, the addition of other cognitive fluency scores to this model did not lead to a significant further explanation of variance in mid-clause pause duration change, $\Delta F(4, 31) = 1.54$, $\Delta R^2 = 7.7\%$, $p = .22$. This finding can potentially be ascribed to the redundant predictors. The regression model based solely on Maze RT (i.e., simple linear regression) was marginally significant and accounted for a similar amount of variance in the fluency change, $F(1, 34) = 4.03$, $\Delta R^2 = 4.9\%$, $p = .053$.

Table 3. Results of Multiple Regression Analyses

[Mid-clause pause frequency]

Predictor	**B**	**95% CI**	***β***	***t***	***p***
Constant	−0.03	[−0.29, 0.24]		−0.21	.84
Mid-clause pause frequency at T1	−0.54	[−0.83, −0.25]	−0.54	−3.80	<.001
Semantic Classification RT	0.00	[0.00, 0.04]	0.04	0.21	.84
Semantic Classification CV	−0.14	[−0.57, 0.29]	−0.12	−0.68	.50
Maze RT	0.00	[0.00, 0.00]	−0.01	−0.06	.95
Maze CV	0.43	[0.04, 0.82]	0.30	2.23	.03

[Mid-clause pause duration]

Predictor	**B**	**95% CI**	***β***	***t***	***p***
Constant	−0.41	[−1.16, 0.34]		−1.12	.27
Mid-clause pause duration at T1	−0.54	[−0.70, −0.39]	−0.83	−6.97	<.001
Semantic Classification RT	0.00	[0.00, 0,00]	−0.10	−0.72	.47
Semantic Classification CV	0.51	[−0.59, 1.60]	0.13	0.94	.35
Maze RT	0.00	[0.00, 0.00]	0.31	2.31	.03
Maze CV	0.59	[−0.53, 1.70]	0.12	1.07	.29

Note. See full regression analyses results in Appendix S4, as well as correlation coefficients between T1 semantic classification and maze tasks and utterance fluency gains in Appendix S5 in the Online Supplementary File.

6. Discussion

The dual purpose of this study was achieved by (a) examining the development of automatization in lexical and syntactic processing (RQ1), and (b) the role of automaticity in lexical and syntactic processing in L2 utterance development (RQ2).

Regarding RQ1, three criteria of automaticity were evaluated. First, RT significantly decreased both for semantic classification (lexical) and maze (syntactic) tasks with substantial effect sizes. Second, CV did not significantly change for either task. Third, meaningful positive correlations between RT and CV were only noted for the semantic classification task at T1 and T2 ($r = .40$ and $r = .29$, respectively), while no such relationship was noted for the maze task at either time point ($r = -.06$ and .07).

These findings, yielded by fine-grained analyses of cognitive fluency (i.e., automatization), suggest distinct developmental trajectories for lexical and grammatical knowledge in a foreign language context. In other words, the array of current evidence indicates higher propensity for automaticity in lexical processing relative to grammatical processing. Specifically, positive RT–CV correlations were found at T1 as well as T2, which is the most important criterion of automatization and/or "restructuring" of some initial routines, such as L1 translation for semantic processing (Segalowitz & Segalowitz, 1993). As this relationship was already established at T1, processing of the lexical items tested in the current study (90% of which were most frequent 1,000 words in English) were presumably automatized sufficiently at the beginning of this study (see also Akamatsu, 2008, who also suggested that highly frequent words were processed automatically, based on the CV results related to Japanese university students). Therefore, the significant decrease in RT from T1 to T2 achieved by this sample of learners may indicate further acceleration of lexical speed after restructuring.

Unlike lexical processing, there was no RT–CV correlation in the maze task at T1 or T2, and the reduction in CV from T1 to T2 was not significant. These findings could potentially be due to the limited learning opportunities allowed in this study for restructuring in syntactic processing (i.e., automatization) to occur. In their study, Suzuki and Sunada (2018) recruited a similar group of L2 English learners at a different Japanese university, and reported a significant RT–CV correlation (i.e., the evidence for automatization of syntactic processing) only in the scores obtained by students who had immersion experience ($M = 15$ months). Similarly, Lim and Godfroid (2015) found a significant RT–CV correlation for their participants, who were recruited in an English as second language context where they used their L2 on a daily basis. Taken together, the current findings suggest that a typical foreign language context like the one in focus of this study (where limited opportunities for L2 learning are available [about 160 hours of learning experience on average over a six-month period], Hanzawa, 2021) does not provide sufficient practice opportunities enough to elicit observable changes in the stability of syntactic processing, i.e., automatization.

To address RQ2, we examined the extent to which automatization in lexical and syntactic processing (i.e., RT and CV) at T1 predicts utterance fluency (speed, breakdown, and repair fluency) changes over the six-month study period. The multiple regression analyses revealed that RT and CV in the maze task were significant predictors of mid-clause pause duration and frequency, respectively. Although the explanatory values of these syntactic indices were not particularly strong (9% and 4.9%), these findings suggest that learners whose syntactic processing is faster and more stable at T1 tended to improve their fluency from T1 to T2. Therefore, the current findings underscore the importance of automatization in grammar processing for developing speaking fluency in a foreign language context over six months.

The current findings contrast the patterns noted by Segalowtiz and Freed (2004) and Leonard and Shea (2017) in the study abroad context. In these two studies, although the research design was similar, the effect of automatization or cognitive fluency on utterance fluency development was inconsistent. Several potential reasons can be offered for the discrepancy, one of which is the difference between study abroad and foreign language contexts. In the study abroad context, ample opportunities to practice speaking L2 are available (Leonard & Shea, 2017) and most learners are able to develop utterance fluency to the extent that their prior L2 processing ability becomes less relevant. In contrast, in the current foreign language context where fewer opportunities to speak L2 were available, grammar processing potentially played a more important role in predicting utterance fluency changes. Therefore, a more direct comparison between at-home and study-abroad context among similar groups of leaners may be useful to further explore this issue.

There are also methodological differences between the current investigation and these previous study-abroad studies. For instance, although Segalowitz and Freed (2004) used both RT and CV, they only included lexical processing task, not grammar task. In contrast, Leonard and Shea (2017) examined both grammar and lexical processing but did not compute CV. This asymmetry highlights th However, as xamining both RT and CV for lexical and grammar processing (and possibly other aspects such as phonological processing) in one study. Although the utility of CV is still under debate (e.g., Hulstijn et al., 2009; Lim & Godfroid, 2015; Suzuki & Sunada, 2018), it is important to keep our minds open and carefully assess its utility as an indicator of automaticity and a predictor of meaningful L2 development. Therefore, we recommend computing CV whenever RT is measured for studying automaticity because "once researchers have obtained RT data, they basically get the CV_{RT} measure for free" (Godfroid, 2019, p. 448).

It should be noted that, in the current study, both RT and CV related to the maze task predicted one specific aspect of utterance fluency, i.e., mid-clause pause changes. As discussed in the Literature Review, cross-sectional research consistently indicates that L2-specific linguistic knowledge and processing efficiency is strongly linked to mid-clause pauses (e.g., Kahng, 2020; Suzuki & Kormos, 2022). The current longitudinal study extends this finding and suggests that L2 processing efficiency (i.e., automatization) can influence not only mid-clause pausing *performance* but also its *change*.

Furthermore, the interdependence between RT and mid-clause pause *duration* and CV and mid-clause pause *frequency* is worth noting. Although there is no prior strong theoretical framework for interpreting duration and frequency, these aspects presumably reflect different cognitive processes. In theory (Segalowitz & Segalowitz, 1993), acceleration of syntactic processing (presumably captured by RT) indicates that pre-established cognitive operations become faster (i.e., it reflects a quantitative change), whereas more stable processing signifies potential restructuring of such mental operations (i.e., reflecting qualitative changes) in the speech processing. Applying this perspective to the current finding, the RT–pause *duration* link may be construed as evidence for greater expediency in syntactic processing (i.e., speed-up or quantitative change), whereas the CV–pause *frequency* link may reflect attainment of processing stability (i.e., restructuring or qualitative change) in syntactic processing. Of course, given the complexities of L2 speech processes, the duration and frequency phenomena can never be "pure" reflections of underlying cognitive fluency. However, our findings point to an interesting pattern, and it may be useful for elaborating the link between the utterance fluency measures and the underlying cognitive process components.

7. Limitations and directions for future research

There are several limitations ~~to~~ in this study that should be noted when interpreting the findings ~~reported in this work.~~ First, the sample size ($n = 39$) was rather small in part due to the technical errors which prevented capturing all data related to the cognitive fluency tasks. This sample size may not be small in light of previous longitudinal study-abroad research that was based on 19 (Segalowitz & Freed, 2004) and 39 (Leonard & Shea, 2017) learners. However, this sample size is inevitably small in this type of research (mainly due to the laborious coding of detailed fluency measures). Therefore, although the coding of detailed fluency measures is laborious, the current findings should be further attested with a larger sample size.

Second, L1 fluency, which influences some aspects of L2 fluency, was not controlled in the current study. As L1 fluency has been found to influence L2 fluency (Kahng, 2020), it is important to tease apart L2-specific fluency components from other general language production processes by controlling L1 speech data.

Third, in the current study, only highly frequent words were utilized to examine the automatization of lexical processing. Because majority of the participants seemed to be able to process those lexical items fairly automatically at the onset of the experiment (i.e., T1) at the group level, it would be interesting to select less frequent lexical items (those not featuring in the 1,000 most frequent word list) that are familiar to them but not processed efficiently in future studies.

Fourth, as pointed out by a reviewer, because individual learners' experience of using English inside and outside the classroom varied, the generalizability of the current finding would require further evidence from different types of L2 learners in different EFL contexts. It might be useful for authors of those studies to report detailed information of learning experience such as the amount of input, teaching methods in English classes, and types of activities, which could be collected through classroom observations and/or questionnaires.

Last but not least, the current six-month longitudinal EFL research revealed limited development in some aspects of cognitive and fluency development (e.g., insubstantial speed fluency development and CV reduction in grammar processing). Because recent EFL classroom research shows that short-term fluency training can result in substantial gains in utterance fluency (e.g., Tran & Saito, 2021; Suzuki & Hanzawa, 2021), it would be worthy of investigating the extent to which such intervention extended over multiple semesters can influence cognitive and utterance fluency in EFL contexts.

8. Conclusions

The aim of this longitudinal study was to investigate the development of automatization in second language (L2) and its role in speaking fluency development in Japan. Based on the fine-grained analyses of cognitive fluency tasks (RT and CV) and speech samples (speed, breakdown, and repair fluency), we demonstrated that L2 learners improved processing speed in their lexical and grammar processing over a six-month period in the EFL context. While no significant improvement was noted for processing stability for either lexical or grammar processing, lexical processing of high-frequency items was found to be relatively stable at the onset of the experiment. Furthermore, the multiple regression analyses revealed that higher levels of automaticity in grammar processing (indicated by faster RT and smaller CV) at T1 significantly predicted the reduction in mid-clause pause duration and frequency. Because the speech breakdown pertaining to mid-clause pauses presumably reflects the degree of L2 linguistic encoding/formulation efficiency, there may be a tight link between automatization in grammar processing and fluency development in an EFL context.

References

Akamatsu, N. (2008). The effects of training on automatization of word recognition in English as a foreign language. *Applied Psycholinguistics*, *29*(2), 175–193.

Allison, P.D. (1990). Change scores as dependent variables in regression analysis. *Sociological Methodology*, *20*, 93–114.

Anderson, J.R., Bothell, D., Byrne, M.D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, *111*(4), 1036–1060.

Boersma, P., & Weenink, D. (2016). Praat: Doing phonetics by computer. Version 6.0.14. Retrieved from http://www.praat.org/

Bosker, H.R., Pinget, A.F., Quene, H., Sanders, T., & de Jong, N.H. (2013). What makes speech sound fluent? The contributions of pauses, speed and repairs. *Language Testing*, *30*(2), 159–175.

Cobb, T. (2002). *Web Vocabprofile* – an adaptation of Heatley, Nation & Coxhead's (2002) *Range*. Retrieved February 2022, from http://www.lextutor.ca/vp/

Collins, L., & Muñoz, C. (2016). The foreign language classroom: Current perspectives and future considerations. *The Modern Language Journal*, *100*(S1), 133–147.

de Jong, N.H., Steinel, M.P., Florijn, A., Schoonen, R.O.B., & Hulstijn, J.H. (2013). Linguistic skills and speaking fluency in a second language. *Applied Psycholinguistics*, *34*(5), 893–916.

de Jong, N.H., & Wempe, T. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods*, *41*(2), 385–390.

DeKeyser, R. (2001). Automaticity and automatization. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp. 125–151). Cambridge, UK: Cambridge University Press.

DeKeyser, R. (2007). *Practice in a second language: Perspectives from applied linguistics and cognitive psychology*. Cambridge, UK: Cambridge University Press.

DeKeyser, R. (2017). Knowledge and skill in ISLA. In S. Loewen & M. Sato (Eds.), *The Routledge handbook of instructed second language acquisition* (pp. 15–32). New York, NY: Routledge.

Elgort, I. (2011). Deliberate learning and vocabulary acquisition in a second language. *Language Learning*, *61*(2), 367–413.

Forster, K., & Forster, J.C. (2003). DMDX: A windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computer*, *35*(1), 116–124.

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics*, *21*(3), 345–375.

Godfroid, A. (2019). Sensitive measures of vocabulary knowledge and processing: Expanding nation's framework. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 433–453). New York, NY: Routledge.

Hanzawa, K. (2021). Development of second language speech fluency in foreign language classrooms: A longitudinal study. *Language Teaching Research*, Advance online publication.

Housen, A., & Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics*, *30*(4), 461–473.

Hui, B., & Godfroid, A. (2020). Testing the role of processing speed and automaticity in second language listening. *Applied Psycholinguistics*, *42*(5), 1089–1115.

Hulstijn, J. H., Van Gelderen, A., & Schoonen, R. O. B. (2009). Automatization in second language acquisition: What does the coefficient of variation tell us? *Applied Psycholinguistics*, *30*(4), 555–582.

Kahng, J. (2014). Exploring utterance and cognitive fluency of L1 and L2 English speakers: Temporal measures and stimulated recall. *Language Learning*, *64*(4), 809–854.

Kahng, J. (2020). Explaining second language utterance fluency: Contribution of cognitive fluency and first language utterance fluency. *Applied Psycholinguistics*, *41*(2), 457–480.

Keith, T. Z. (2015). *Multiple regression and beyond: An introduction to multiple regression and structural equation modelling* (2nd ed.). New York, NY: Routledge.

Kormos, J. (2006). *Speech Production and Second Language Acquisition*. London, UK: Routledge.

Kormos, J., & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System*, *32*(2), 145–164.

Lambert, C., Aubrey, S., & Leeming, P. (2020). Task preparation and second language speech production. *TESOL Quarterly*, *55*(2), 331–365.

Leonard, K. R., & Shea, C. (2017). L2 speaking development during study abroad: Fluency, accuracy, complexity, and underlying cognitive factors. *The Modern Language Journal*, *101*(1), 1–15.

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

Lim, H., & Godfroid, A. (2015). Automatization in second language sentence processing: A partial, conceptual replication of Hulstijn, Van Gelderen, and Schoonen's 2009 study. *Applied Psycholinguistics*, *36*(5), 1247–1282.

Lyster, R., & Ballinger, S. (2011). Content-based language teaching: Convergent concerns across divergent contexts. *Language Teaching Research*, *15*(3), 279–288.

McManus, K., & Marsden, E. (2018). Signatures of automaticity during practice: Explicit instruction about L1 processing routines can improve L2 grammatical processing. *Applied Psycholinguistics*, *40*(1), 205–234.

Moors, A. (2016). Automaticity: Componential, causal, and mechanistic explanations. *Annual Review of Psychology*, *67*, 263–287.

Phillips, N. A., Segalowitz, N., O'Brien, I., & Yamasaki, N. (2004). Semantic priming in a first and second language: Evidence from reaction time variability and event-related brain potentials. *Journal of Neurolinguistics*, *17*(2–3), 237–262.

Pili-Moss, D., Brill-Schuetz, K. A., Faretta-Stutenberg, M., & Morgan-Short, K. (2019). Contributions of declarative and procedural memory to accuracy and automatization during second language practice. *Bilingualism: Language and Cognition*, *23*(3), 639–651.

doi Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning*, *64*(4), 878–912.

doi Saito, K., Ilkan, M., Magne, V., Tran, M. N., & Suzuki, S. (2018). Acoustic characteristics and learner profiles of low-, mid- and high-level second language fluency. *Applied Psycholinguistics*, *39*(3), 593–617.

doi Saito, K., Suzukida, Y., & Sun, H. (2018). Aptitude, experience and second language pronunciation proficiency development in classroom settings: A longitudinal study. *Studies in Second Language Acquisition*, *41*(1), 201–225.

doi Tran, M. N., & Saito, K. (2021). Effects of the 4/3/2 activity revisited: Extending Boers (2014) and Thai & Boers (2016). *Language Teaching Research, Early View*.

doi Segalowitz, N. (2003). Automaticity and second languages. In C. J. Doughty & M. H. Long (Eds.), *The Handbook of Second Language Acquisition* (pp. 392–408). Hoboken, NJ: Blackwell Publishing.

doi Segalowitz, N. (2010). *Cognitive bases of second language fluency*. New York, NY: Routledge.

doi Segalowitz, N. (2016). Second language fluency and its underlying cognitive and social determinants. *International Review of Applied Linguistics in Language Teaching*, *54*(2), 75–95.

doi Segalowitz, N., & Freed, B. F. (2004). Context, contact, and cognition in oral fluency acquisition: Learning Spanish in at home and study abroad contexts. *Studies in Second Language Acquisition*, *26*, 173–199.

doi Segalowitz, N., & Segalowitz, S. (1993). Skilled performance, practice, and the differentiation of speed-up from automatization effects: Evidence from second language word recognition. *Applied Psycholinguistics*, *14*(3), 369–385.

doi Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review*, *84*(2), 127–190.

doi Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, *30*(4), 510–532.

doi Spada, N. (1997). Form-focussed instruction and second language acquisition: A review of classroom and laboratory research. *Language Teaching*, *30*(2), 73–87.

doi Suzuki, S., & Kormos, J. (2020). Linguistic dimensions of comprehensibility and fluency: An investigation of complexity, accuracy, and fluency in second language argumentative speech. *Studies in Second Language Acquisition*, *42*(1), 143–167.

doi Suzuki, S., & Kormos, J. (2022). The multidimensionality of second language oral fluency: Interfacing cognitive fluency and utterance fluency. *Studies in Second Language Acquisition*. Advance online publication.

doi Suzuki, Y. (2018). The role of procedural learning ability in automatization of L2 morphology under different learning schedules. *Studies in Second Language Acquisition*, *40*(4), 923–937.

doi Suzuki, Y., & Hanzawa, K. (2021). Massed task repetition is a double-edged sword for fluency development: An EFL classroom study. *Studies in Second Language Acquisition*, 1–26.

doi Suzuki, Y., & Sunada, M. (2018). Automatization in second language sentence processing: Relationship between elicited imitation and maze tasks. *Bilingualism: Language and Cognition*, *21*(1), 32–46.

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). Amsterdam, The Netherlands: John Benjamins.

Online supporting information

Appendix S1

The results of Wilcoxon signed-rank test for utterance fluency change between Time 1 and Time 2

Appendix S2

The number of participants identified as outliers for each fluency measure

Appendix S3

Correlations between T1 utterance fluency measures and fluency gains

Appendix S4

Hierarchical multiple regression results

Appendix S5

Partial correlations between T1 cognitive fluency measures and fluency gains

Address for correspondence

Yuichi Suzuki
Cross-Cultural and Japanese Studies
Kanagawa University
3-27-1 Rokkakubashi
Kanagawa-ku, Yokohama-shi, Kanagawa 221-8686
Japan
yuicchi0819@gmail.com
https://orcid.org/0000-0002-1197-0315

Co-author information

Keiko Hanzawa
Institute of Arts and Sciences
Tokyo University of Science
k.hanzawa@rs.tus.ac.jp

Publication history

Date received: 22 April 2022
Date accepted: 6 March 2023

Author Queries

- Please provide a shorter running header for this article.
- Please provide missing content for Appendix S1, S2, S3, S4, and S5 as referenced in this article.