A case study: longitudinal development of oral fluency in production of monologic narratives among two adult Japanese learners of English.

Teruaki Tsushima

ABSTRACT

The present article reports a case study designed to examine development of oral fluency in two adult Japanese learners of English (P1, P2) during a relatively long learning period (one year, and one and half years, respectively). The two participants regularly made oral recordings of monologic narratives primarily about their daily lives. Among the recorded narratives, eight narratives per month were subjected to acoustic analyses (N=96, N=144, respectively) on the measures of fluency, which included the speed and composite measures (e.g., articulation rate, speech rate, mean length of runs), the breakdown measures (e.g., frequency and duration of between-clause and within -clause pauses), and the repair measures (e.g., repetition, self-correction). Data from native speakers of English (N=15) were also analyzed.

The results showed that the speed and composite measures significantly improved toward the native means both in P1 and P2. It was found, however, that the improvement was not linear as there existed a series of sub-periods where there was little improvement. Second, while the frequency of between-clause pauses did not substantially decline in P1, it decreased significantly from the middle of the learning period in P2. On the other hand, the frequency of within-clause pauses significantly declined from the earliest sub-periods in both participants, indicating that the frequency of within-clause pauses declined in the earlier sub-period than that of between-clause pauses. It was also found that, when the between-clause pauses were broken down into clause-final and clause-initial pauses, the significant decline in the between-clause pauses was largely due to clause-initial, rather than clause-final pauses. Analyses on the frequency of repairs showed that P2 made a larger number of repairs than P1 even though the proficiency level was higher, supporting the hypothesis that the repair measures

do not distinguish different proficiency levels. Implications for teaching were also discussed.

key words; fluency, pause, L2 learning, speech production model

1. Introduction

1-1. Review of literature

Most foreign language instructors may agree that one of the important aims of foreign language instruction is to improve oral fluency. As oral fluency significantly reflects learners' language proficiency and communicative ability, it is used as one of the essential constructs in L2 (second language) assessment (e.g., IELTS[1], TOEFL iBT[2], CEFR[3]). Historically, oral fluency has been defined in a number of ways. Recently, Tavacoli and Hunter (2018) proposed considering fluency as a pyramid made up of four levels. At the very broad sense, fluency is defined as a general view of second language proficiency, which include L2 ability in skills beyond L2 speaking. At the next, broad sense, fluency refers to competent L2 speaking ability. At the narrow perspective, it refers to ease, flow, and continuity of speech without regard to grammatical complexity and accuracy. Finally, at the very narrow perspective, fluency is defined in terms of its measurable features such as speed, breakdown, and repair. Generally speaking, research on fluency focuses on the very narrow sense of fluency.

Segalowits (2010, pp. 46-52) proposed a model of fluency which expanded our understanding of how fluency could be conceptualized. According to the model, fluency comprises three interrelated domains; "Cognitive fluency refers to the efficiency of the speaker's underlying processes responsible for fluency-relevant features of utterances, utterance fluency refers to the oral features of utterances that reflect the operation of underlying cognitive processes, and perceived fluency refers to the inferences that listeners make about a speaker's cognitive fluency based on perception of the utterance fluency features of the speaker's speech output (p. 50)." In the previous research, utterance fluency has been examined to make inferences about speakers' cognitive fluency. The utterance fluency is made up of three components: speed fluency (e.g., articulation rate), breakdown fluency (e.g., pause frequency), and repair fluency (e.g., self-repairs). Following the previous research, the present study examined the utterance fluency, attempting to examine how the speed, breakdown, and repair fluency improved in two adult Japanese

learners of English during a relatively long period of L2 learning (i.e., one year, and one and a half years).

Previous research has attempted to interpret the findings of the utterance fluency in the theoretical framework of a modular model of speech production in L1 (the first language) proposed by Levelt (1989), and later adapted to L2 speech production by Kormos (2006). These models postulate three stages of speech production: conceptualization, formulation, and articulation stage. In the conceptualization stage, a desired concept or intention is generated, and an intended message called a preverbal plan is formulated. In the formulation stage, the preverbal plan results in a phrase structure through a series of processes, including lexical access/retrieval and syntactic/morpho-phonological encoding processes. On the phrase structure, phonetic and phonological information is encoded, which results in a phonetic plan, which consists of articulatory scores which specify how articulatory gestures are activated. In the articulation stage, the articulatory scores generate articulatory movements of the vocal apparatus (e.g., tongue, lips).

The model postulates an important distinction between serial and parallel processing. For native speakers and proficient L2 speakers, lexical access/retrieval and syntactic/morphological encoding processing in the formulation stage are assumed to operate simultaneously with the other stages (i.e., parallel processing). For less proficient L2 speakers, mental operations, especially lexical access/retrieval and the encoding processes in the formulation stage, require a great deal of attentional resources, so that one module cannot operate simultaneously with the other ones (i.e., serial processing). This limitation is expected to produce more fluency breakdown in the speech of less proficient L2 speakers. The relation between components of the utterance fluency (i.e., speed, breakdown, repair) and the postulated stages of speech production described above may not be straightforward. It has been suggested, however, that the speed fluency is associated with all the stages, and that between-clause pauses (i.e., breakdown) are primarily associated with the conceptualization stage, while within-clause pauses with the formulation stage (Lambert, Aubrey, & Leeming, 2020; Lambert, Kormos, & Minn, 2017).

Previous research has investigated how the measures in the utterance fluency are related to development of L2 proficiency (Baker-Smemoe, Dewey, Brown, & Martinsen, 2014; Ginther, Dimova, & Yang, 2010; Iwashita, Brown, McNamara, & O'Hagan, 2008; Saito, Ilkan, Magne, Tran, & Suzuki, 2018; Parvaneh Tavakoli, 2010; P. Tavakoli, Nakatsuhara, & Hunter, 2020). Saito et al. (2018), for example, examined how the speed, breakdown, repair measures of spontaneous speech samples were related to perceived levels (i.

e., low, mid, and high) of oral proficiency among 90 adult Japanese learners of English with diverse second language experience. It was found that the speed measure (i.e., articulation rate) distinguished all the levels, showing consistent increase among participants from the low to high level. It was also found that the frequency of final-clause pauses (i.e., the breakdown) distinguished low- and mid-level fluency performance, and the frequency of the mid-clause pauses differentiated all the performance levels. Repair measures (i.e., repetition ratio and self-correction ratio) did not produce consistent results. More recently, Tavakoli et al. (2020) examined which of speed, breakdown, repair, and composite measures characterized fluency at assessed levels of proficiency, and which consistently distinguished one level from the next. It examined fluency in 32 speakers performing 4 tasks of the British Council's Aptis Speaking Test, who then were divided into four different levels of proficiency according to CEFR (i.e., A2, B1, B2, C1). It was found that speed and composite measures (e.g., speech rate) distinguished fluency from the lowest to upper-intermediate levels (A2 to B2). It was also found that frequency of mid-clause silent pauses (breakdown) distinguished lower levels of proficiency (A2 and B1) from higher levels (B2 and C1), indicating that frequency of mid-clause silent pauses decreased with an increase in proficiency. On the other hand, the number of end-clause silent pauses (breakdown) did not distinguish the different levels. Consistent with the result of Saito et al. (2018), repair measures did not consistently differentiate the different levels.

In sum, the available evidence suggests that speed and composite measures improve across all the levels of proficiency, and that frequency of mid-clause silent pauses also improve across the levels of proficiency although there might exist some non-linearity in the way it decreases. The two studies provided inconsistent result regarding frequency of final-clause pauses. In Saito et al. (2018), it distinguished the low from the mid-level group, while in Tavakoli et al. (2020), it did not distinguish different levels. Finally, it suggests that repair measures may not be significantly related to improvement across proficiencies.

Previous research has provided longitudinal (as opposed to cross-sectional) data on development of fluency. Vallas-Ferrer & Mora (2014), for example, examined development of the utterance fluency among 27 Catalan/Spanish learners of English who participated in 6 months of formal instruction (FI), followed by 3 months of study abroad (SA), and by another 6 months of FI. It was found that four out of six fluency measures (i.e., speech rate, mean length of runs, phonation-time ratio, and pause duration ratio) significantly improved during SA, but not in FI before or after SA. The results suggested that it might be difficult to improve L2 learners' fluency without intensive and extensive

speech practice afforded by the SA context. More recently, Tsushima（2019, 2020）conducted a series of case studies in which multiple fluency measures（e.g., articulation rate, speech rate, mean length of runs）in production of monologic narratives by adult Japanese learners of English were longitudinally examined for a relatively long period of time（e.g., for 12 to 24 months）. The participants received weekly individual-based speech training in addition to the English classes in the FI context. It was found that most of the fluency measures significantly improved during the training period. In addition, Pipe and Tsushima（2021a, 2021b）examined the utterance fluency of 12 adult Japanese learners of English who received "Timed-Pair Practice" which was designed to promote L2 learners' fluency. During the course of one year where data were obtained ten times, the speed measure（i.e., articulation rate）and the composite measures（i.e., speech rate, mean length of runs）significantly improved. The results suggested that, even in the FI context, L2 earners were able to improve fluency if they received a special method of speech training.

1-2.　Rationale

　　Although previous research has examined various aspects of fluency development in L2 learning, there is little longitudinal data in which fluency data, especially breakdown and repair measures, are obtained periodically with relatively short intervals for a long period of time. Although the previous research might suggest that some fluency measures（e.g., articulation rate, mid-cause pauses）may linearly improve across the proficiency levels, some non-linear developmental pattern might be detected in the longitudinal data. In addition, some interactions or relations among the measures may be found as well. Another gap in the literature is that there is little data available on fluency development which focused on the lower levels of L2 learning. More specifically, the previous research focused on comparing the performances across the CEFR levels of A2 and above. However, a great deal of developmental changes in fluency are expected to take place *within* the A2 level, where learners' productions consist of a great number of pauses and hesitations. Finally, more detailed analyses of pauses are warranted to better understand the learning process of fluency. For example, previous research has made a distinction between the final-clause pauses and mid-clause pauses. However, the mid-clause pauses may be further divided into the between-phrase pauses and the within-phrase pauses.

　　To fill in these gaps, the present study examined fluency development of two participants（P1 & P2）who were at the CEFR level of A2 at the entry of the study for a period of one year（P1）and one and a half years（P2）, respectively. Their monologic spontane-

ous productions were obtained eight times a month, totaling 96 and 144 data sets, respectively. The fluency measures included the speed measures (e.g., articulation rate), composite measures (e.g., speech rate, mean length of runs), breakdown measures (e.g., frequency and length of between-clause pauses and mid-clause pauses), and repair measures (e.g., self-correction). The mid-clause pauses were further divided into sub-categories such as between-phrase pauses and within-phrase pauses.

1-3. Specific Research Questions

Specific questions asked in the present study were the following: 1) How did the speed and composite fluency improve in P1 and P2 during the course of L2 learning? 2) How did breakdown fluency, especially the frequency and duration of between-clause and within-clause pauses improve? 3) How did the repair fluency improve?

2. Method

2-1. Participants

The participants (P1, P2) were two female students at a private university in Tokyo, Japan. They were monolingual speakers of Japanese who had never lived in an English-speaking country. P1 joined the research program in July of her freshman year and has been in the program for approximately one year. During the freshman year, she attended general English classes twice a week, which focused on English conversation. She then joined an academic program called a global career program from April of her sophomore year. During the first semester of the sophomore year, she attended English conversation classes every weekday, which were taught by native instructors. As assessed by standard speaking tests (i.e., CASEC[4] speaking test) and her TOEIC scores, her English proficiency level in terms of CEFR was mid-A2 at the entry of the program and high-A2 at the end.

P2 joined the research program in February of her freshman year and has been in it for approximately one and half years. She joined an academic program called the advanced English program from April of her sophomore year. During the sophomore year and the first semester of her junior year, she attended English classes twice a week; one class focused on written English (i.e., reading, writing, grammar) taught by Japanese instructors, and the other on oral English (i.e., speaking) taught by native instructors. Also assessed by the standard speaking tests and her TOEIC scores, her English proficiency

level in terms of CEFR was high-A2 at the entry of the program and lower-B1 at the end. Both of them were highly motivated to study English, especially to improve their speaking skills.

1-1.　Data acquisition

Speech data were based on the participants' one-minute monologic narrative productions about a topic that included daily events, class activities, memories of a trip, hobbies, an opinion on a familiar issue and others. The participants were asked to make a recording every day, but at least eight times a month. P1 made 362 recordings during the course of one year (i.e., almost every day), while P2 made 166 recordings during the course of one and a half years (i.e., 9.2 recordings per month on the average). The author made comments about their recordings in terms of pronunciation, grammatical mistakes, and choice of expressions, which were sent to the participants via email.

They made recordings in a quiet place at their homes, using an i-phone with a high-quality microphone (*Zoom IQ7*) with a pop filter attached to it. For recording, *Zoom Handy Recorder App* was used with a sampling rate of 44,000 Hz and 16 bits of resolution. The sound file was saved in a wav format and sent to the author via email, then low-pass filtered at 8,000 Hz and normalized for average intensity at 70 *dB* on sound analysis software, *Praat* (Boersma & Weenink, 2014). On *Praat*, the sound waves were segmented at syllable boundaries, and repairs (i.e., false starts, self-corrections, and repetitions) were coded.

Speech data of native speakers of English were obtained in a quiet environment at the university of Edinburgh, England. The participants (*N*=15) were native speakers of English from England, U.S.A. or Canada, and were asked to talk about their life events for about one minute. The sound files were processed in the same way as the Japanese participants.

1-2.　Speech training

The participants received a weekly speech training (about 90 minutes each) conducted by the author. It primarily focused on instruction on pronunciation, but also included speaking practice, instruction on vocabulary, and reading practice. P1 received the training from the beginning until the end of her freshman year (i.e., for 9 months), while P2 for the whole period (i.e., for one and a half years). Overall, 26 and 65 practice sessions were held, respectively. It should be noted that the main aim of the training was improve-

ment of pronunciation, especially segmental and prosodic production, rather than that of fluency such that the effect of the training on the observed fluency development might be indirect at best.

1−3. Analysis procedure

The present study used the speed and composite measures, the breakdown measures, and the repair measures as described below. A pause was defined as a silent period of 250 ms or above, following previous research （e.g., Saito et al., 2018; P. Tavakoli et al., 2020）.

1） Speed measures

・AR （Articulation Rate）: the total number of words produced in a narrative divided by the amount of time taken to produce it （excluding pause time） expressed in minutes.

2） Composite measures

・SR （Speech Rate）: the total number of syllables produced in a narrative divided by the amount of total time required to produce it （including pause time） expressed in minutes.

・MLoR （Mean Length of Runs）: the average number of syllables produced in utterances between pauses of 250 ms and above.

3） Breakdown measures

・PauseRat （Pause Duration Ratio）: the total length of pauses divided by the total amount of speaking time （including pause time）.

・PauseFreqBC （Frequency of Between-Clause Pauses per 100 syllables）: the number of between-clause pauses divided by the number of syllables times 100. Example: I live in Tokyo // and // I study at a university. This includes all the pauses that take place from the end of the clause and the beginning of the next clause that starts with its subject. This is broken down into PauseFreqCF and PauseFreqIF below.

・PauseDurBC （Duration of Between-Clause Pauses）: the average length of between-clause pauses in seconds.

・PauseFreqCF （Frequency of Clause-Final Pauses per 100 syllables）: the number of clause-final pauses divided by the number of syllables times 100. This includes a pause after the end of a clause. Example: I live in Tokyo // and I study at a university.

・PauseFreqCI （Frequency of Clause-Initial Pauses per 100 syllables）: the number of clause-initial pauses divided by the number of syllables times 100. This includes pauses that take place after a clause final pause and the beginning of the next clause that

starts with its subject. Example: I live in Tokyo because // I study at a university there.

・PauseFreqWC（Frequency of Within-Clause Pauses per 100 syllables）: the number of within-clause pauses divided by the number of syllables times 100. This is broken down into PauseFreqWCPB and PauseFreqWCWPB below. Example: I live // in Tokyo.

・PauseDurWC（Duration of Within-Clause Pauses）: the average length of within-clause pauses in seconds.

・PauseFreqWCPB（Frequency of Within-Clause, Phrasal Boundary Pauses）per 100 syllables）: the number of within-clause, phrasal boundary pauses divided by the number of syllables times 100. Example: I live // in the outskirts // of Tokyo.

・PauseDurWCPB（Frequency of Within-Clause, Phrasal Boundary Pauses）per 100 syllables）: the average length of within-clause, phrasal boundary pauses.

・PauseFreqWCWPB（Frequency of Within-Clause, Within-Phrasal Boundary Pauses）per 100 syllables）: the number of within-clause, within-phrasal boundary pauses divided by the number of syllables times 100. Example: I live in the outskirts of // Tokyo.

・PauseDurWCWPB（Frequency of Within-Clause, Phrasal Boundary Pauses）per 100 syllables）: the average length of within-clause, within-phrasal boundary pauses.

4）Repair measures

・FreqRep（Frequency of Repairs）: the number of repairs divided by the number of syllables times 100. Repairs included false starts, repetitions, and self-corrections.

For P1, the total period（12 months）was divided into four sub-periods（i.e., T1, T2, T3, and T4）, while for P2, the total period（18 months）was divided into six sub-periods（i.e., T1, T2, T3, T4, T5, and T6）. For each month, eight recordings were randomly selected for analyses, which resulted in 24 speech samples per sub-period（i.e., 3 months）.

3. Results

3-1. Speed and composite measures

Table 1 shows the speed measure（AR）and the composite measures（SR, MLoR）across the sub-periods in P1 and P2. Those of the native data are also shown. Statistical analyses used Kruskal-Wallis H tests（i.e., non-parametric tests）to test the significance of overall changes across the sub-periods, and non-parametric, Mann-Whitney U tests（i.e., non-parametric tests）to test the significance of changes in the adjacent sub-periods. In order to protect against an inflated type 1 error, the significance level was set at $p=0.01$.

	Sub-Period		T1	T2	T3	T4	T5	T6	NS
P1	AR	M	146.9	150.1	167.3	182.6			267.5
		SD	13.9	12.4	11.0	13.9			30.8
	SR	M	68.1	72.7	87.4	105.9			227.7
		SD	11.9	10.1	8.2	12.6			33.6
	MLoR	M	2.7	2.8	3.4	4.1			16.9
		SD	0.5	0.5	0.5	0.6			5.5
P2	AR	M	167.2	170.8	174.8	198.9	196.5	209.6	
		SD	14.8	13.1	11.3	17.4	14.2	14.7	
	SR	M	89.9	104.7	116.4	130.8	133.0	144.3	
		SD	16.9	11.7	11.9	14.2	12.4	12.6	
	MLoR	M	4.1	4.4	4.9	5.9	5.7	6.4	
		SD	1.0	0.6	0.8	1.1	1.0	1.0	

Table 1. Mean (M) and one standard deviation (SD) of the speed and composite measures averaged over narratives recorded within each sub-period. P1/P2=Japanese participants; NS=native speakers of English; SR=speech rate; AR=articulation rate; MLoR=mean length of runs.

In both participants, all the measures increased toward the NS (native speakers') means across the sub-periods. In P1, none of the changes was significant between T1 and T2, while all the changes were significant between T2 and T3 (AR: U=81.0, z=−4.03, p<0.001, r=0.59; SR: U=66.0, z=−4.36, p<0.001, r=0.64, MLoR: U=90.5, z=−3.82, p<0.001, r=0.56)[5], and T3 and T4 (AR: U=80.0, z=−2.97, p<0.001, r=0.44; SR: U=45.0, z=−3.97, p<0.001, r=0.58, MLoR: U=76.0, z=−3.08, p=0.002, r=0.45). The results indicated that, in P1, the speed and composite measures did not improve for the first six months and improved significantly afterwards. Such non-linear changes were also observed in P2. All the measures improved gradually from T1 to T3 but improved substantially between T3 and T4. However, none of the measures improved between T4 and T5, but improved substantially again between T5 and T6. The changes between T1 and T3 were significant for SR, U=46.0, z=−4.99, p<0.001, r=0.72, and MLoR, U=149.5, z=−2.86, p<0.001, r=0.41, and marginally significant for AR, U=187.0, z=−2.08, p=0.037, r=0.30. None of the changes between T4 and T5 were significant, while all the changes between T3 and T4 (AR: U=63.0, z=−4.53, p<0.001, r=0.66; SR: U=107.0, z=−3.60, p<0.001, r=0.52, MLoR: U=119.0, z=−3.34, p=0.001, r=0.49), and T5 and T6 were significant (AR: U=152.0, z=−2.80, p=0.005, r=0.40; SR: U=143.0, z=−2.98, p=0.003, r=0.43, MLoR: U=161.0, z=−2.61, p=0.001, r=0.38). The results indicated that, in P2, there were periods of substantial improvement (i.e., between T3 and T4, and between T5 and T6), as well as little improvement which might be called a plateau (i.e., between T4 and T5).

It should be noted that, reflecting the difference in the proficiency level at the entry

of the learning period（i.e., low A2 for P1 and upper A2 for P2）, all the measures were higher in P2 than in P1 at T1. Also reflecting the difference in the proficiency level at the end of the learning period（i.e., upper A2 for P1, and low B1 for P2）, all the measures were higher in P2 than in P1. In addition, NS means of all the measures were substantially higher than those of the last learning period（i.e., T4 for P1, and T6 for P2）, indicating that there was still a lot of room for improvement for both P1 and P2. In sum, the overall data indicated that both P1 and P2 became able to speak with greater speed（AR, SR）, and produce a greater number of syllables between pauses（MLoR）through the learning period, although the way the abilities improved was not linear with some plateaus where there was little improvement for an extended period of time.

3-2.　Breakdown measures（Pause Ratio）

As is shown in Table 2, the pause ratio changed toward the NS mean across the sub-periods in both P1 and P2. The pause ratio in P1 did not significantly change between T1 and T2. The change between T2 and T3 was marginally significant, U=161.0, z=-2.27, p=0.023, r=0.33, and was significant between T3 and T4, U=75.0, z=-3.11, p=0.002, r=0.46. The result indicated, as was observed in the speed and composite measures, the pause ratio started to improve after T3. In P2, the pause ratio decreased significantly between T1 and T2, U=137.0, z=-3.26, p=0.001, r=0.47, and T2 and T3, U=122.0, z=-3.56, p<0.001, r=0.51, and did not change significantly between T3 and T6. The result indicated that, in P2, the amount of pause time in one narrative did not change substantially after T3, even though the speed of speaking markedly improved. This finding suggested that P2 needed the approximately same amount of pause time even when she spoke with greater speed. The overall results showed that both P1 and P2 became able to significantly reduce the amount of pause time in one narrative during the learning periods.

	Sub-Period		T1	T2	T3	T4	T5	T6	NS
P1	PauseRat	M	54.0	52.0	48.0	42.0			15.0
		SD	6.0	5.0	5.0	6.0			5.0
P2	PauseRat	M	46.0	39.0	33.0	34.0	32.0	31.0	
		SD	8.0	4.0	5.0	4.0	4.0	3.0	

Table 2. Mean（M）and one standard deviation（SD）of the breakdown measure（Pause Ratio） averaged over narratives recorded within each sub-period. P1/P2=Japanese participants; NS=native speakers of English; PauseRat=pause ratio.

	Sub-Period		T1	T2	T3	T4	T5	T6	NS
P1	PauseFreqBC	M	19.1	20.6	18.5	16.6			5.3
		SD	4.5	3.6	2.7	2.7			1.9
	PauseDurBC	M	1.41	1.43	1.30	1.02			0.57
		SD	0.27	0.30	0.24	0.13			0.19
P2	PauseFreqBC	M	16.8	16.2	14.9	12.9	12.6	11.7	
		SD	4.9	2.2	2.9	2.8	2.2	2.2	
	PauseDurBC	M	1.42	1.04	0.89	0.96	0.88	0.89	
		SD	0.44	0.21	0.09	0.14	0.09	0.11	

Table 3. Mean (M) and one standard deviation (SD) of the breakdown measures (frequency and duration of between-clause pauses) averaged over narratives recorded within each sub-period. P1/P2=Japanese participants; NS=native speakers of English; PauseFreqBC=frequency of between-clause pauses; PauseDurBC=duration of between-clause pauses.

3-3. Breakdown measures (Frequency and duration of between-clause pauses)

Table 3 shows that the frequency and duration of between-clause pauses decreased toward the native means in both P1 and P2. In P1, the overall decrease in the frequency was marginally significant, $H(3)=10.95$, $p=0.012$, $\eta^2=0.09$[6], while the decrease between adjacent sub-periods was not, indicating that the number of between-clause pauses did not substantially decrease through the sub-periods. The overall decrease in duration of between-clause pauses in P1, on the other hand, was significant, $H(3)=28.62$, $p<0.001$, $\eta^2=0.31$, mostly due to the significant decrease between T3 and T4, $U=40.0$, $z=-4.11$, $p<0.001$, $r=0.59$. In P2, the overall decrease in frequency of between-clause pauses was significant, $H(5)=48.87$, $p<0.001$, $\eta^2=0.32$. In particular, the decrease was not significant between T1 and T2, while it was significant between T2 and T4, $U=95.0$, $z=-3.90$, $p<0.001$, $r=0.56$. The overall decrease in duration of between-clause pauses was also significant, $H(5)=61.71$, $p<0.001$, $\eta^2=0.41$. The decrease was significant between T1 and T2, $U=99.0$, $z=-4.02$, $p<0.001$, $r=0.57$, and T2 and T3, $U=133.0$, $z=-3.34$, $p=0.001$, $r=0.48$, but it remained at the same level after T4. The overall results found that, in P1, the frequency and duration of the between-clause pauses showed a sign of decline at T4, indicating that she was able to reduce the number and duration of the between-clause pauses at the end of the learning period. P2 was able to reduce the number of the between-clause pauses through the learning period, while the duration hit the floor in the middle of the learning period (i.e., T3). This finding suggested that the frequency and duration did not necessarily improve in the same way. Especially, the duration appeared to have some floor effect where it was difficult to further decrease below a certain level.

3-4.　Breakdown measures（Frequency of clause-final and clause-initial pauses）

As described above, the between-clause pauses analyzed in 3-3 were broken down into clause-final and clause-initial pauses. First of all, as is shown in Table4, it is noticeable that the duration of clause-final pauses（PauseDurCF）was longer than that of clause-initial pauses（PauseDurCI）in both P1 and P2. This might be due to the fact that clause-final pauses occur after completion of the previous sentence or clause such that the speaker might be engaged in generating new concepts or intentions, formulating a preverbal message for the following sentence（i.e., conceptualization stage）. On the other hand, clause-initial pauses appear to function as a way of providing the speaker with supplementary time to further organize the preverbal message.

In P1, the magnitude of the decline in either the frequency of clause-final or of clause-initial pauses was not so substantial, and the overall change was only marginally significant（$H(3)=7.66$, $p=0.054$, $\eta^2=0.06$, $H(3)=8.27$, $p=0.041$, $\eta^2=0.06$, respectively）. On the other hand, the overall decline in the duration of the clause-final pauses was significant, $H(3)=25.23$, $p<0.001$, $\eta^2=0.26$, which was mostly due to the significant decline between T3 and T4, $U=64.0$, $z=-3.43$, $p=0.001$, $r=0.54$. The overall decline in the duration of the clause-initial pauses was also significant, $H(3)=14.14$, $p=0.003$, $\eta^2=0.26$, indicating that both duration of the clause-final and clause-initial pauses became significantly shorter across the sub-periods. In P2, the overall decline was significant in the frequency of the clause-final pauses, $H(5)=19.15$, $p=0.002$, $\eta^2=0.10$, as well as that of the clause initial pauses, $H(5)=27.3$, $p<0.001$, $\eta^2=0.16$. However, the magnitude of the decrease was much larger in the frequency of clause-initial than that of clause-final pauses. Specifically, while the frequency of clause-final pauses stopped decreasing after T4, the frequency of clause-initial pauses continued to decline throughout the sub-periods. The results indicated that the significant decline in the frequency of the between-clause pauses（see 3-3）was largely due to that of the clause-initial pauses rather than the clause-final pauses. The overall decrease in the duration of the clause-final pauses was significant, $H(5)=36.11$, $p<0.001$, $\eta^2=0.22$, mostly due to the decline between T1 and T3, $U=53.0$, $z=-4.85$, $p<0.001$, $r=0.69$. The overall decrease in the duration of the clause-initial pauses was also significant, $H(5)=59.63$, $p<0.001$, $\eta^2=0.39$, indicating that, as was observed in P1, the duration of both clause-final and clause-initial pauses became significantly shorter through the sub-periods. Overall, the finding that the duration of both clause-final and clause initial pauses significantly decreased in both participants suggested that they became more efficient in constructing the preverbal message, and at the same time, retrieving/accessing the lexical items, and encoding the mor-

	Sub-Period		T1	T2	T3	T4	T5	T6	NS
P1	PauseFreqCF	*M*	8.1	9.4	8.7	7.4			2.4
		SD	2.5	2.2	1.6	1.7			1.0
	PauseDurCF	*M*	1.71	1.87	1.62	1.25			0.62
		SD	0.36	0.45	0.32	0.22			0.22
	PauseFreqCI	*M*	11.0	11.3	9.8	9.2			3.0
		SD	3.4	2.6	2.2	1.6			1.5
	PauseDurCI	*M*	1.17	1.08	1.03	0.84			0.52
		SD	0.32	0.26	0.30	0.17			0.22
P2	PauseFreqCF	*M*	6.9	7.4	6.7	6.0	6.0	5.9	
		SD	2.2	1.8	1.5	1.1	1.3	1.2	
	PauseDurCF	*M*	1.92	1.27	1.11	1.18	1.13	1.17	
		SD	0.81	0.30	0.15	0.22	0.18	0.19	
	PauseFreqCI	*M*	9.9	8.7	8.2	7.0	6.6	5.8	
		SD	5.0	2.4	2.3	2.9	1.6	2.1	
	PauseDurCI	*M*	1.06	0.86	0.70	0.73	0.64	0.59	
		SD	0.27	0.23	0.11	0.18	0.12	0.12	

Table 4. Mean (*M*) and one standard deviation (*SD*) of the breakdown measures (frequency of clause-final and clause-initial pauses) averaged over narratives recorded within each sub-period. P1/P2=Japanese participants; NS=native speakers of English; PauseFreq (Dur) CF=frequency (duration) of clause-final pauses; PauseFreq (Dur) CI=frequency (duration) of clause-initial pauses.

pho-syntactic information of the following clause.

3-5. Breakdown measures (Frequency and duration of within-clause pauses)

Table 5 shows that, in both P1 and P2, the frequency and duration of within-clause pauses decreased toward the NS means. In P1, the overall decrease of the frequency of within-clause pauses was significant, $H(3)=48.68$, $p<0.001$, $\eta^2=0.54$. The decrease was significant between T1 and T2, $U=149.0$, $z=-3.01$, $p=0.003$, $r=0.43$, T2 and T3, $U=111.5$, $z=-3.36$, $p=0.001$, $r=0.50$, and marginally significant between T3 and T4, $U=109.5$, $z=-2.13$, $p=0.033$, $r=0.34$. The overall decrease of the duration of within-clause pauses was significant, $H(3)=15.74$, $p=0.001$, $\eta^2=0.15$. The decrease was significant between T1 and T2, $U=164.0$, $z=-2.71$, $p=0.007$, $r=0.38$, but the duration appeared to stabilize after T2. The results indicated that, unlike the speed and composite measures and the frequency and duration of between-clause pauses, the frequency and duration of within-clause pauses decreased from the beginning of the sub-periods. In P2, the overall decrease of the frequency of within-clause pauses was significant, $H(5)=31.30$, $p<0.001$, $\eta^2=0.19$. The decrease was significant between T1 and T3, $U=134.5$, $z=-3.17$, $p=0.002$, $r=0.47$, but the frequency fluctuated around the same level after T4. The overall decrease of the duration of within-clause pauses was also significant, $H(5)=42.05$, $p<0.001$, $\eta^2=0.27$. The decrease was significant

	Sub-Period		T1	T2	T3	T4	T5	T6	NS
P1	PauseFreqWC	M	23.9	18.3	12.6	10.0			2.1
		SD	6.2	5.6	3.6	5.0			1.4
	PauseDurWC	M	0.92	0.77	0.74	0.72			0.52
		SD	0.22	0.16	0.13	0.15			0.13
P2	PauseFreqWC	M	10.3	7.2	6.5	5.0	5.8	4.8	
		SD	4.5	3.1	2.7	2.5	3.2	2.2	
	PauseDurWC	M	0.92	0.82	0.66	0.69	0.64	0.55	
		SD	0.27	0.16	0.19	0.27	0.17	0.12	

Table 5. Mean（M）and one standard deviation（SD）of the breakdown measures（frequency and duration of within-clause pauses）averaged over narratives recorded within each sub-period. P1/P2=Japanese participants; NS=native speakers of English; PauseFreqWC=frequency of within-clause pauses; PauseDurWC=duration of within-clause pauses.

between T1 and T3, U=107.0, z=−3.73, p<0.001, r=0.54, but it stopped decreasing at T3. The results indicated that, as was observed in P1, the frequency of within-clause pauses started to decline in the earlier sub-period than the frequency of between-clause pauses. In sum, the overall findings indicated that both P1 and P2 became able to reduce the frequency and duration of between-clause pauses during the learning period, suggesting that they became more efficient in access/retrieval of lexical items and encoding of morphosyntactic information.

3−6. Breakdown measures（Frequency and duration of within-clause, phrasal pauses）

To conduct finer analyses of the within-clause pauses presented in 3−5, they were divided into the following two categories: within-clause, phrasal pauses（PauseFreq（Dur）WCPB, presented in 3−6）and within-clause, within-phrasal boundary pauses（PauseFreq（Dur）WCWPB, presented in 3−7）. The within-clause, phrasal pauses occur at a phrasal boundary（e.g., I lived // in Kobe // many years ago）, while the within-clause, within-phrasal boundary pauses occur within a phrasal boundary（e.g., I lived in // Kobe many // years ago）.

As shown in Table 6, the frequency and duration of the within-clause, phrasal pauses decreased toward the NS means in both P1 and P2, although the frequency was much larger in P1 than in P2. In P1, the overall decrease in the frequency across the sub-periods was significant, $H(3)$=37.70, p<0.001, η^2=0.41, and the decrease was significant between T2 and T3, U=117.0, z=−3.24, p=0.001, r=0.47, while it was marginally significant between T1 and T2, U=202.5, z=−1.93, p=0.053, r=0.28, and T3 and T4, U=99.0, z=−2.43, p=0.015, r=0.39. The overall decrease in the duration across the sub-periods was also significant, H

	Sub-Period		T1	T2	T3	T4	T5	T6	NS
P1	PauseFreqWCPB	*M*	18.2	15.0	10.8	8.2			1.5
		SD	5.6	4.5	3.2	3.7			1.0
	PauseDurWCPB	*M*	0.98	0.79	0.78	0.71			0.59
		SD	0.27	0.15	0.16	0.16			0.16
P2	PauseFreqWCPB	*M*	8.6	6.1	5.5	4.0	5.2	4.2	
		SD	3.9	2.8	2.1	2.1	3.2	2.2	
	PauseDurWCPB	*M*	0.95	0.85	0.69	0.76	0.64	0.59	
		SD	0.28	0.22	0.22	0.31	0.20	0.14	

Table 6. Mean (M) and one standard deviation (SD) of the breakdown measures (frequency and duration of within-clause pauses) averaged over narratives recorded within each sub-period. P1/P2=Japanese participants; NS=native speakers of English; PauseFreqWCPB=frequency of within-clause, phrasal pauses; PauseDurWCPB=duration of within-clause, phrasal pauses.

(3)=17.06, p=0.001, η^2=0.16, and the decline was significant between T1 and T2, U=157.0, z=−2.85, p=0.004, r=0.41. However, the duration stayed approximately at the same level after T2. The results indicated that, in P1, the frequency of the within-clause, phrasal pauses decreased across the sub-periods, while the duration decreased at the initial sub-periods.

In P2, the overall decrease in the frequency of the within-clause, phrasal pauses was significant, $H(5)$=28.84, p<0.001, η^2=0.17, and the decrease was significant between T1 and T4, U=71.0, z=−4.36, p<0.001, r=0.64. However, it fluctuated around the same level between T4 and T6. As for the duration, the overall decrease was also significant, $H(5)$=32.03, p<0.001, η^2=0.19, and the decrease was significant between T1 and T3, U=123.0, z=−3.40, p=0.001, r=0.49. Nevertheless, the duration remained around the same level after T3. The results indicated that, in P2, both the frequency and duration of the within-clause, phrasal pauses declined in the earlier subperiods. The overall results indicated that both P1 and P2 were able to reduce the frequency and duration of the within-clause, phrasal pauses from the beginning of the learning period.

3−7. Breakdown measures (Frequency and duration of within-clause, within-phrasal pauses)

Table 7 shows that, in both P1 and P2, the frequency of within-clause, within-phrasal pauses declined toward the NS means across the sub-periods. In P1, the overall decrement was significant, $H(3)$=23.16, p<0.001, η^2=0.24, with the decrease being significant between T1 and T2, U=167.5, z=−2.64, p=0.008, r=0.38, and marginally significant between T2 and T3, U=184.5, z=−1.77, p=0.077, r=0.26. The duration fluctuated around the same region

	Sub-Period		T1	T2	T3	T4	T5	T6	NS
P1	PauseFreqWCWPB	M	5.7	3.3	1.9	1.9			0.6
		SD	3.4	3.1	1.7	1.9			0.6
	PauseDurWCWPB	M	0.75	0.71	0.66	0.85			0.54
		SD	0.27	0.27	0.35	0.51			0.18
P2	PauseFreqWCWPB	M	1.7	1.1	1.0	1.0	0.6	0.6	
		SD	1.7	1.3	1.3	1.0	0.6	0.7	
	PauseDurWCWPB	M	0.87	0.84	0.57	0.54	0.58	0.55	
		SD	0.71	0.48	0.21	0.19	0.31	0.25	

Table 7. Mean (M) and one standard deviation (SD) of the breakdown measures (frequency and duration of within-clause pauses) averaged over narratives recorded within each sub-period. P1/P2=Japanese participants; NS=native speakers of English; PauseFreqWCWPB=frequency of within-clause, within-phrasal pauses; PauseDurWCWPB=duration of within-clause, within-phrasal pauses.

across the sub-periods. In P2, the frequency was much lower than P1 at T1, and reached the native level at T5, with the overall decrease being marginally significant, $H(5)=10.72$, $p=0.057$, $\eta^2=0.04$. The duration reached the native level at T3, with the overall change being not significant in P2. The results indicated that the frequency of within-clause, within-phrasal pauses declined up to T3 in P1, and that P2 made relatively few within-clause, within-phrasal pauses from the beginning of the sub-periods.

3-8.　Repair measures (Frequency of repairs)

As shown in Table 8, the frequency of repairs in P1 was at the native level at T1. It even decreased significantly to a much lower level at T2, $U=145.0$, $z=-3.63$, $p<0.001$, $r=0.52$, indicating that P1 rarely made use of repairs after T2. In P2, the frequency of repairs declined substantially between T1 and T2 and continue to gradually decrease toward the NS mean between T2 and T6. The overall decrease was significant, $H(5)=40.43$, $p<0.001$, $\eta^2=0.26$, with the decrease between T1 and T2 being significant, $U=121.0$, $z=-3.58$, $p<0.001$, $r=0.51$. The results indicated that P2 was able to reduce the frequency of repairs at the beginning of the sub-periods but continued to produce a certain number of repairs throughout the sub-periods.

4.　Discussion and Conclusion

4-1.　Overall improvement

The present study was designed to examine whether and how the fluency measures (i.e., speed and composite measures, breakdown measures, and repair measures) im-

	Sub-Period		T1	T2	T3	T4	T5	T6	NS
P1	FreqRepair	M	1.73	0.13	0.10	0.24			1.81
		SD	2.14	0.43	0.35	0.43			1.67
P2	FreqRepair	M	9.94	5.75	4.94	4.42	4.64	3.57	
		SD	4.48	2.53	2.52	2.32	1.68	1.57	

Table 8. Mean (*M*) and one standard deviation (*SD*) of the repair measures (frequency of repairs) averaged over narratives recorded within each sub-period. P1/P2=Japanese participants; NS=native speakers of English; FreqRepair=frequency of repairs.

proved in two adult Japanese learners of English during a relatively long learning periods (i.e., one year, and one and half years, respectively). It was found that all the fluency measures significantly changed toward the NS means both in P1 and P2. The results clearly demonstrated that both participants became more fluent, speaking with a greater speed, producing a larger chunk of syllables, reducing the frequency and duration of between-clause and within-clause pauses, and reducing the number of repairs. For P1, the improvement might be due to the speech training, general English classes she was taking, and the speaking activities associated with making recordings every day. The observed significant improvement between T3 and T4, in particular, might be brought about by the intensive speaking practice in the global career program during the first semester of the sophomore year. For P2, the improvement might be attributable to the speech training, the advanced English classes she was taking, and the speaking activities associated with making recordings. The results supported the hypothesis that it is possible to improve L2 fluency under the formal instruction settings if the learners are given adequate opportunities to regularly practice speaking L2.

4-2. Improvement of the speed and composite measures

Despite the overall significant improvement of fluency, it was found that the improvement was not always linear, and what might be called "plateaus" may exist during the course of the fluency development. In P1, for example, the speed and composite measures did not significantly increase until T2, but significantly improved from T2 through T4. The results suggested that P1 was not able to speak with a significantly greater speed and with larger chunks of syllables during the first half of the sub-periods. In P2, there was a set of periods when little increase was observed in terms of the speed and composite measures as well as the frequency of non-clausal boundaries (i.e., T4 and T5), while there were a set of periods when the degree of increase in terms of the speed and composite measures was much larger than the others (i.e., T3 and T4, T5 and T6). This pat-

tern of development might fit into "the plateau effects" reported in the areas of skill acquisition (e.g., Rechards, 2008). The results suggest that it may take a relatively long period of time (i.e., three to six months) of input and output practice to go beyond the current level of fluency. Previous research has suggested that, in term of the production models described above (Kormos, 2006), the speed and composite measures are associated with all the stages of the production process (i.e., conceptualization, formulation, and articulation stage). The existence of the plateau might suggest that, in the formulation stage in particular, it may take a great amount of experience and practice to speed up the process of lexical access/retrieval and syntactic/morpho-phonological encoding processes. It might be the case that speeding up such a process requires a large number of attempts to utilize a number of lexical items and syntactic-morphological structures.

4-3.　Improvement of the breakdown measures (within-clause and between-clause pauses)

As for the breakdown measures of within-clause and between-clause pauses, the present results supported the hypothesis that the decrease in the frequency of the within-clause pauses may begin earlier than that of the between-clause pauses. In P1, for example, while the frequency of between-clause pauses did not show significant decrease, that of within-clause pauses showed significant decrease from T1, indicating that P1 was able to reduce the number of within-clause pauses from the beginning through the end of the sub-periods. In P2, the frequency of between-clause pauses significantly decreased from T2 and the following sub-periods, while the frequency of within-clause pauses significantly decreased from T1 up to T4 and stayed at the same level until the end of the sub-periods. It might be the case that, at the earlier phases of L2 learning, the participants start to reduce pauses in the boundaries of smaller syntactic units. This hypothesis is supported by the results of the analyses on within-clause, phrasal pauses, and that of within-clause, within-phrasal pauses. The frequency of within-clause, within-phrasal pauses became reduced to the lower level earlier than the within-clause, phrasal pauses in both P1 and P2. This suggested that the participants might first attempt to make an uninterrupted chunk of a relatively basic syntactic unit (e.g., a noun phrase).

4-4.　Improvement of the breakdown measures (clause-final and clause-initial pauses)

Another interesting finding of the present study was that, in P2, when the between-clause pauses were further divided into clause-final and clause-initial pauses, the overall

significant decrease was largely due to the decrease in the frequency of the clause-initial pauses. This might reflect the tendency for the participant to use conjunctions such as" and" and "because" between the clauses. These pauses might be used between these conjunctions and the beginning of the following clause to construct a preverbal message and at the same time, lay out a syntactic and phonological structure of the following sentence. In this sense, it is plausible that the clause-initial pauses might be associated with the conceptualization stage of the production model. The reduction of frequency and duration of the clause-initial pauses, thus, suggests that the processes involved in this stage become speedier and more efficient. The present results suggested the importance of making a distinction between the clause-final and clause-initial pauses, as most previous research (e.g., Saito et al., 2018; P. Tavakoli et al., 2020) either analyzed only clause-final pauses or mid-clause pauses.

4-5. Improvement of the repair measures

As for the repair measures, the present results were compatible with the hypothesis that repairs do not reliably distinguish between proficiency levels (Saito et al., 2018; P. Tavakoli et al., 2020). It was found that P1 rarely made repairs although she made a relatively great number of between- and within-clause pauses. On the other hand, P2 made use of a relatively larger number of repairs, especially in the early sub-periods, although her proficiency level was higher than that of P1. Specifically, she used 7.52 repairs per 100 syllables in the first sub-periods. It appears that making use of repairs in the narrative production might be a matter of individual tendency or choice rather than that of common developmental pattern.

4-6. Relation between fluency measures and proficiency levels

What do the present results suggest about the relation between the fluency measures and proficiency levels? As mentioned above, P1's proficiency level in terms of CEFR was from mid-A2 to high-A2, while that of P2's was from high-A2 to low-B1. Previous research provided a hypothesis that the speed and composite measures would significantly improve between A2 and B1 (Saito et al., 2018; P. Tavakoli et al., 2020), and the present results were compatible with this hypothesis as they showed significant improvement in the speed and composite measures both in P1 and in P2. As for the breakdown measures, previous research provided a hypothesis that the mid-clause pauses might improve between A2 and B1 level, and this hypothesis was consonant with the present results, as

both P1 and P2 showed significant reduction in the frequency and duration of the mid-clause pauses. With regard to the clause-final pauses, previous research provided conflicting hypotheses (Saito et al., 2018; P. Tavakoli et al., 2020), namely the frequency of clause-final pauses would, or would not significantly decrease between A2 and B1. The present results were more in agreement with the latter hypothesis as the frequency of clause-final pauses did not significantly decrease either in P1 or P2. However, given that the frequency in the final sub-period was much higher than that of the native speakers, it might be possible that it would further decrease toward the NS means as their proficiency improves in future.

4−7.　Teaching implications

The present study implies that a monologic narrative production task might be an effective method of practice to improve fluency. According to DeKeyser (2007), improvement of fluency involves automatization of syntactic, morphological, and phonetic encoding processes and the use of prefabricated language units called formulaic language. In order to facilitate the automatization processes and the use of formulaic language, it is recommended that L2 learners be provided with the opportunity to practice repeating a set of syntactic and morpho-phonological rules (i.e., procedural knowledge of the encoding processes). It is also recommended that they should practice under more and more demanding conditions where they are required to express real ideas and intentions. In the monologic narrative production task, the learners are provided with a great deal of opportunities to utilize a certain set of syntactic and morpho-phonological rules, while gaining the procedural knowledge of the encoding processes. They are also required to express a variety of ideas and intentions in real life. In addition, the speaking time of one minute appears to be a good choice for the learners at the CEFER level of A2 and B1. The speakers can usually talk about one event from the beginning to its ending in around one minute. Thus, it is recommended that L2 instructors use the monologic narrative production task to improve students' fluency in the speaking classes or as homework assignments.

Second, the present results suggest that it may take a relatively long time to improve fluency, especially in the formal instruction settings where exposure to the target language is limited. In P1, the composite measures (i.e., SR and MLoR) started to improve substantially after a plateau of about half a year. P2 also showed a plateau of about half a year in terms of the composite measures, preceded and followed by a period of substantial improvement. Similar results were reported in the author's previous similar case study

with the learners in the study-abroad settings (Tsushima, 2019), and in the formal instruction settings (Tsushima, 2020). It is recommended that instructors show understanding about the period of little improvement and encourage the learners to continue practicing until the next period of substantial improvement.

4−8. Limitations of the present study

The present study was limited in terms of the following points. First of all, as the case study with only two participants, it is not clear whether or to what extent the present results were generalizable to other populations. Second, the present study lacks the data on perceived fluency. Namely, no data are available as to how native speakers of English evaluate the degree of fluency of the produced narratives. Additional research is certainly needed to investigate this point. Finally, the present data are based on a relatively short narrative task (i.e., approximately one minute), which is repeated a great number of times. It might be the case that the participants became highly skilled in this kind of task, and the results might not be generalizable to other speaking tasks.

4−9. Concluding remarks

Even with these limitations, the present study has shown that the two participants were able to significantly improve their fluency during the learning period. One of the advantages of a case study is that it can provide hypotheses to be tested in a larger study. On this point, there is an ongoing study with a larger pool of participants ($N=13$), which uses a similar data acquisition and data analysis procedure. The currently available results based on four months of data are highly compatible with the present results. The study is expected to provide further data on the hypotheses discussed in the present study.

Acknowledgement

Note────────

1）The International English Language Testing System

2）Test of English as a Foreign Language, internet-Based Test

3）Common European Framework of Reference for Languages

4）Computerized Assessment System for English Communication

5）According to Cohen（1988）, the effect size is small if the value of r varies around 0.1, medium if it varies around 0.3, and large if it varies around 0.5.

6）According to Cohen（1988）, the effect size is small if the value of η^2 varies around 0.01, medium if it varies around 0.06, and large if it varies around 0.14.

References

Baker-Smemoe, W., Dewey, D. P., Brown, J., & Martinsen, R. A.（2014）. Variables affecting L2 gains during study abroad. *Foreign Language Annals, 47*, 464–486.

Cohen, J.（1988）. *Statistical Power Analysis for the Behavioral Sciences*（2nd ed.）. Hillsdale, NJ: Erlbaum.

DeKeyser, R. M.（2007）. *Practice in a Second Language : Perspectives from Applied Linguistics and Cognitive Psychology*. New York: Campridge University Press.

Ginther, A., Dimova, S., & Yang, R.（2010）. Conceptual and empirical relationships between temporal measures of fluency and oral English proficiency with implications for automated scoring. *Language Testing, 27*, 379–399.

Iwashita, N., Brown, A., McNamara, T., & O'Hagan, S.（2008）. Assessed levels of second language speaking proficiency: How distinct? *Applied Linguistics, 29*, 24–49.

Kormos, J.（2006）. *Speech Production and Second Language Acquisition*. Mahwah, N.J.: Lawrence Erlbaum Associates.

Lambert, C., Aubrey, S., & Leeming, P.（2020）. Task preparation and second language speech production. *TESOL Quarterly*. doi: 10.1002/tesq.598

Lambert, C., Kormos, J., & Minn, D.（2017）. Task repetition and second language processing. *Studies in Second Language Acquisition, 39*, 167–196.

Levelt, E. J. M.（1989）. *Speaking ; from Intention to Artculation*. Cambridge: A bradford book: the MIT press.

Pipe, J., & Tsushima, T.（2021a）. The application of suprasegmental features of pronunciation into the classroom through the timed-pair-practice framework. *Journal of Humanities & Natural Sciences, 148*, 31–70.

Pipe, J., & Tsushima, T.（2021b）. *Improved fluency through the Timed-pair-practice framework*. Paper presented at the Asian Conference on Language 2021, Tokyo.

Rechards, J. C.（2008）. *Moving Beyond the Plateau: From Intermediate to Advanced Levels in Language Learning*. New York: Cambridge University Press.

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*, 593–617.

Segalowitz, N.（2010）. *Cognitive Bases of Second Language Fluency*. New York : Routledge.

Tavakoli, P.（2010）. Pausing patterns : differences between L2 learners and native speakers. *ELT Journal, 65*（1）, 71-79.

Tavakoli, P., & Hunter, A. M.（2018）. Is fluency being 'neglected' in the classroom ? Teacher understanding of fluency and related classroom practices. *Language Teaching Research, 22*（3）, 330-349.

Tavakoli, P., Nakatsuhara, F., & Hunter, A. M.（2020）. Aspects of fluency across assessed levels of speaking proficiency. *The Modern Language Journal, 104*（1）, 169-191. doi: 10.1111/modl.12620

Tsushima, T.（2018）. An exploratory case study : improvement of the ability to produce the rhythmic properties of English in a spontaneous speech task through long-term, individual-based speech training. *Journal of Humanities & Natural Sciences, 143*, 15-41.

Tsushima, T.（2019）. An exploratory case study on development of fluency and English speech rhythm through individual speech training and a study abroad program. *Journal of Humanities & Natural Sciences, 145*, 29-63.

Tsushima, T.（2020）. Development of fluency and English speech rhythm through individual-based, long-term speech training in the formal instruction settings. *Journal of Humanities & Natural Sciences, 147*, 59-93.

Valls-Ferrer, M., & Mora, J. C.（2014）. L2 fluency development in formal instruction and study abroad : The role of initial fluency level and language contact. In C. Pérez-Vidal（Ed.）, *Language Acquisition in Study Abroad and Formal Instruction Contexts*（pp. 111-136）. Amsterdam : John Benjamins Publishing Company.