An interim report of a case study: longitudinal development of L2 oral fluency during a study abroad program

Teruaki Tsushima

ABSTRACT

The present article is an interim report on the ongoing study which investigates longitudinal development of L2 (second language) oral fluency in an adult learner of English (P1) who belongs to a study abroad program. The present article reports the data up to approximately one and a half months into the study abroad period (SA, henceforth). The data were based on monologic spontaneous narratives regularly recorded by P1 for two and a half years (N=270). Among them, 196 narratives were submitted to acoustic analyses. The analyses examined 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 found that P1 significantly improved the speed and composite measures through the pre-SA period, although she showed a period of apparent regression in the middle. Second, although the frequency of between-clause pauses significantly decreased during the pre-SA period, the decrease was largely due to that of clause-initial pauses rather than clause-final pauses. Third, the frequency of within-clause pauses also significantly decreased during the pre-SA period, while it declined in the earlier stage of development than that of between-clause pauses. Fourth, the overall frequency of repairs significantly decreased during the pre-SA period. Finally, P1 has shown a sign of further improvement in a number of fluency measures during less than two months of SA. Implications for teaching were also discussed.

key words; L2 speech learning, fluency, speed, pause, repair, study abroad

1. Introduction

1-1. Review of literature

It is generally believed that the study abroad (SA, henceforth) settings provide second language (L2, henceforth) learners with an advantageous learning environment where they have an ample exposure to the target language and have a great deal of opportunities to interact with native as well as nonnative speakers from different countries both inside and outside of the language classes. On the other hand, in the formal instruction (FI, henceforth) settings, the L2 learners generally do not have a sufficient exposure to the target language and lack adequate opportunities to use it outside the classrooms. For this reason, many educational institutions offer their students a variety of SA programs. A private university which the author is affiliated with has an SA program called Global Career Program (GCP, henceforth), which is a three-year program which consists of the pre-SA period (Pre-SA, henceforth) in Japan, an approximately five-month SA period in Australia, and the post-SA period (Post-SA, henceforth) in Japan. The present paper is an interim report of an ongoing case study where linguistic development of one student in GCP is recorded and analyzed from the beginning to the end of the program. Among the linguistic abilities, the present study focused on development of English speech fluency in spontaneous narrative productions.

It is generally agreed that one of the major objectives of language learning for L2 learners is to achieve speech fluency. Accordingly, it is one of the primary pedagogic goals of many foreign language teaching programs. In addition, speech fluency is used as a prime construct in the assessment of L2 oral language proficiency (e.g., IELTS[1], TOEFL iBT[2], CEFR[3]). Speech fluency in L2 research generally refers to ease, flow, and continuity of speech without regard to grammatical complexity and accuracy and is commonly measured in terms of such features as speed, breakdown, and repair (Tavakoli & Hunter, 2018). According to Segalowits (2010), fluency is composed of cognitive fluency, utterance fluency, and perceived fluency. The present study focused on the utterance fluency and measured speed fluency (e.g., articulation rate), composite fluency (e.g., speech rate), breakdown fluency (e.g., pause frequency), and repair fluency (e.g., self-repairs).

Following previous research on L2 fluency, the observed developmental changes in the learners' fluency were interpreted in the theoretical framework of L1 (the first

language) speech production proposed by Levelt (1989), which was adapted to L2 speech production by Kormos (2006). In these models, three independent modules (i.e., conceptualizer, formulator, and articulator) are postulated. At the conceptualization module, for example, a speaker generates a desired concept or intention, and formulates an intended message called a preverbal plan. At the formulator module, the speaker formulates a phrase structure through lexical access/retrieval and syntactic/morpho-phonological encoding processes. Then, the speaker formulates a phonetic plan through phonetic and phonological encoding processes. At the articulator module, the speaker generates articulatory movements of the vocal apparatus. The model also postulates that the speaker monitors the produced utterance in terms of various aspects of the syntactic, morpho-phonological, and phonetic information, and gives feedback through the external monitoring system.

The previous research (Lambert, Aubrey, & Leeming, 2021; Lambert, Kormos, & Minn, 2017) has suggested that, among the utterance fluency (i.e., speed, breakdown, and repair), the speed fluency (e.g., articulation rate) may reflect all the modules, as it is associated with the efficiency of creating a concept, accessing, and retrieving lexical items, formulating a syntactic and phonological structure, and executing articulation. It has also been suggested that, among the breakdown fluency, between-clause pauses are primarily related with the conceptualizer module because that is where the speaker formulates the intended message of the following utterance (Butterworth, 1975; Götz, 2013; Lambert et al., 2017). On the other hand, it has been suggested that within-clause pauses are primarily associated with the formulator module because they are assumed to reflect the lexical access/retrieval and the syntactic and phonological encoding processes of the following phrase (Götz, 2013; Kormos, 2006; Skehan, Xiaoyue, Qian, & Wang, 2012).

Previous research has examined how the utterance fluency (e.g., speed, breakdown, and repair) is related with the developmental levels 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; Tavakoli, 2010; Tavakoli, Nakatsuhara, & Hunter, 2020). It has been suggested that the speed fluency increases across all the levels of proficiency in terms of a standardized oral proficiency test (Saito et al., 2018; Tavakoli et al., 2020). Regarding the breakdown fluency, it has been suggested that the frequency of mid-clause silent pauses decreases with an increase in L2 proficiency (Saito et al., 2018; Tavakoli et al., 2020), while whether the frequency of end-clause silent pauses distinguishes different levels of proficiency remains inconclusive

(Saito et al., 2018; Tavakoli et al., 2020). Finally, it has been suggested that the repair fluency is not significantly related with the different levels of L2 proficiency (Saito et al., 2018; Tavakoli et al., 2020).

Previous research on the effects of SA on development of fluency conducted a longitudinal study in which the same participants of the SA program were tested before, during, and after SA (Freed, 1995; Lennon, 1990; Towell, Hawkins, & Bazergui, 1996; Valls-Ferrer & Mora, 2014). The previous findings indicated that SA has significant effects on the L2 learners' improvement of fluency, while FI's effect on fluency is fairly limited (Freed, 1995; Lennon, 1990; Towell et al., 1996; Valls-Ferrer & Mora, 2014). Recent studies, however, have indicated that L2 learners' fluency can be improved when they are provided with instruction that utilizes some effective teaching techniques (Pipe & Tsushima, 2021a, 2021b, 2022a, 2022b, 2022 (in press)). A series of studies conducted by Pipe & Tsushima (2021a, 2021b, 2022a, 2022b, 2022 (in press)), for example, examined development of fluency during the first year of GCP in the FI settings using the "Timed-Pair-Practice Framework (TPPF, henceforth)". The studies examined the speed fluency (e.g., articulation rate), composite fluency (e.g., speech rate, mean length of run), breakdown fluency (e.g., pause duration ration, frequency, and duration of between- and within-clause pauses) as well as repair fluency (e.g., self-correction) in the students' spontaneous narrative productions. Although the number of participants was relatively small (e.g., N=11 in Pipe & Tsushima, 2022a), the results showed that all the fluency measures, except for the duration of within-clause pauses and repair fluency, significantly improved during the first two semesters of GCP.

Another line of longitudinal case studies have examined the effects of FI and SA on development of fluency (Tsushima, 2018, 2019, 2020, 2021). Tsushima (2018, 2019), for example, examined the improvement of fluency and speech rhythm among two Japanese learners of English (J1, J2) who were enrolled in GCP. They recorded spontaneously produced narratives (N=272 for J1, and N=284 for J2) over a period of 24 months and 20 months, respectively. The results showed that all the fluency measures significantly improved during Pre-SA and continued to significantly increase during SA in J2. The findings suggested that, under Pre-SA, fluency can be improved through a combination of narrative production practice, individual speech production training, and intensive speaking classes.

1-2. Rationale

The present study attempted to extend the previous studies on the effects of Pre-SA and SA on development of L2 fluency. Although many previous studies investigated the effect of the SA programs using averaged group data (Freed, 1995; Lennon, 1990; Towell et al., 1996; Valls-Ferrer & Mora, 2014), there is little data available on a case study where detailed analyses of individual learners were conducted on development of fluency over a long period of time. In L2 longitudinal research, it is a rule rather than an exception to find individual variability in the way L2 leaners improve their linguistic abilities (e.g., Tsushima, 2008). In development of fluency, for example, the timing and length of so-called plateau and spurt of improvement may greatly differ among individual learners. It was hoped that providing detailed analyses of individual development would detect developmental patterns that might be overlooked in the averaged group data.

The spontaneous narrative production task was used because the speaker was expected to engage online in 1) formulation of an intended message, 2) lexical retrieval/access, 3) phonological, morphological, and syntactic encoding, 4) execution of articulatory movements, and 5) monitoring of their output. Facilitation of processing in each of these production stages was expected to result in improvement in various aspects of the utterance fluency in the task.

Following Tsushima (2019), the present study used the speed fluency measure (i.e., articulation rate), the composite fluency measures (i.e., speech rate and mean length of runs), and the breakdown measures (i.e., pause-phonation ratio). In addition to these measures, the present study conducted more detailed analyses of the breakdown measures. It included the frequency and length of between-clause pauses and within-clause pauses. The within-clause pauses were further divided into sub-categories such as between-phrase pauses and within-phrase pauses (see the method section below for further details). Finally, it also included the repair fluency measures (i.e., false starts, self-correction, repeats).

1-3. Specific research questions

Specific research questions asked in this paper were the following.

1)　How did the fluency measures change in the FI settings before SA and during about a month and a few weeks under the SA settings?

2) How did different fluency measures (i.e., the speed, composite, breakdown, repair) interact with each other?

2. Method

2-1. Participant

The participant (P1) was one female student who had belonged to GCP since the beginning of their freshman year at a private university in Tokyo. P1 had participated in a short-term study abroad program when she was a high school student. However, she can be considered as a monolingual speaker of Japanese as their primary language for daily communication has been dominantly Japanese. In GCP, students normally go to Australia and stay there for approximately five months in the fourth semester of the program (i.e., the second semester of the sophomore year). For P1, however, the study abroad was delayed for one year due to the COVID-19 pandemic. As a result, she did the study abroad in the sixth semester of the program (i.e., the second semester of the junior year). As assessed by a standard speaking test (i.e., Versant) and the TOEIC scores, the English proficiency level in terms of CEFR was low-A2 at the entry program and B1 just before SA. She is highly motivated to study English, especially to improve their speaking and pronunciation skills.

2-2. Data acquisition

Speech data were elicited by asking the participant to produce a monologic narrative for approximately one minute about what had happened in their daily lives. For example, the topic included daily events, class activities, memories of a trip, hobbies, and others. For P1, the data were based on 270 narrative productions over the course of approximately 28 months (i.e., 2 years and 4 months). The first set of 11 productions were recorded weekly in the first semester of the freshmen year during the GCP class which focused on speaking. The rest of the productions (N=259) were recorded at her home from the beginning of August 2020 to the middle of September 2022. P1 was asked to make a recording at least eight times a month and made approximately 9.5 recordings a month. Among the recordings, the first set of 11 productions and the additional 199 productions (about eight productions per month) were analyzed.

In the GCP class, the recordings were made in a quiet environment using a high-quality microphone (*AT4040*) attached to a PCM recorder (*DR-44WL*). In her home,

the recordings were made in a quiet environment, using an iPhone with a high-quality microphone (*Zoom IQ7*) with a pop filter attached to it. For the recording, speech recording software (*Zoom Handy Recorder*) was used with a sampling rate of 44,000 Hz and 16 bits of resolution. The recorded file was sent to the author via email, then, denoised, low pass filtered at 8,000 Hz, and normalized for average intensity at 70 *dB* on sound analysis software, *Praat* (Boersma & Weenink, 2014). Using a *Praat* function (i.e., Annotate to TextGrid (Silences)), sound waves were first segmented between the pause and sound portions. Following the previous research, the pause was defined as a silent period of 250ms and longer (e.g., Saito et al., 2018; Tavakoli et al., 2020). Then, syllable boundaries were manually segmented. Finally, the types of pauses and repairs (see below for their descriptions) were coded. Recordings from native speakers of English were made 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 in a quiet environment. The sound files were analyzed in the same way described just above.

2-3. Speech Training

P1 received a weekly, approximately 90-minute speech training from the beginning of August 2020 to the middle of July 2022 (approximately 2 years) totaling 82 practice sessions. The primary focus of the training was on pronunciation, but it also included various activities including speaking, vocabulary, and reading practice. To prepare for SA, it especially focused on improving speaking skills during the 1st semester of the junior year. Overall, 82 practice sessions were held.

2-4. Analysis procedure

For the analysis of the pauses and repairs, the following measures were used.

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 250ms and above.

3) Breakdown measures

- PauseRat (Pause-Phonation Ratio): the total length of pauses divided by the total amount of speaking time (including pause time).
- Between-clause pauses (PauseBC, henceforth): 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.
  - ○ PauseFreqBC (frequency of PauseBC per 100 syllables)
  - ○ PauseDurBC (duration of PauseBC in seconds)

  PauseBC is broken down into PauseCF and PauseCI below.
- Clause-final pauses (PauseCF, henceforth): This includes a pause after the end of a clause. Example: I live in Tokyo//and I study at a university.
  - ○ PauseFreqCF (frequency of PauceCF per 100 syllables)
  - ○ PauseDurCF (duration of PauceCF in seconds)
- Clause-initial pauses (PauseCI, henceforth): This includes pauses that take place after the 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.
  - ○ PauseFreqCI (frequency of PauseCI per 100 syllables)
  - ○ PauseDurCI (duration of PauseCI in seconds)
- Within-clause pauses (PauseWC, henceforth)
  - ○ PauseFreqWC (frequency of PauseWC per 100 syllables)
  - ○ PauseDurWC (duration of PauseWC)

  PauseWC is broken down into PauseWCPB and PauseWCWPB below.
- Within-clause, phrasal boundary pauses (PauseWCPB, henceforth)
  - ○ PauseFreqWCPB (frequency of PauseWCPB per 100 syllables)
  - ○ PauseDurWCPB (duration of PauseWCPB in seconds)
- Within-clause, within-phrasal boundary pauses (PauseWCWPB, henceforth)
  - ○ PauseFreqWCWPB (frequency of PauseWCWPB per 100 syllables)
  - ○ PauseDurWCWPB (th duration of PauseWCWPB)

4) Repair measures

- FreqRep (frequency of repairs per 100 syllables): the sum of FSFreq, RPFreq, and SCFreq below
- FSFreq (frequency of false starts per 100 syllables): Example: **I am living. I have**

been living in Tokyo since I was three.

- RPFreq（frequency of repeats per 100 syllables）: Example: I am **studying, studying** Korean.
- SCFreq（frequency of self-correction per 100 syllables）: Example: I went to **a market, a supermarket**.

5）Others

- FLFreq（the frequency of fillers per 100 syllables）: Example: I major in, **uh**, economics.

The total period was divided into ten subperiods as follows.

- 1yr1Sem（the 1st semester of the 1st year）
- 1yrSumV（the summer vacation of the 1st year）
- 1yr2Sem（the 2nd semester of the 1st year）
- 1yrSprV（the spring vacation between the 1st and the 2nd year）
- 2yr1Sem（the 1st semester of the 2nd year）
- 2yrSumV（the summer vacation of the 2nd year）
- 2yr2Sem（the 2nd semester of the 2nd year）
- 2yrSprV（the spring vacation between the 2nd year and the 3rd year）
- 3yr1Sem（the 1st semester of the 3rd year）
- SA（the study abroad）

To test the significance of the overall changes across the subperiods, non-parametric Kruskal-Wallis H tests were used. The post-hoc tests used non-parametric, Mann-Whitney U tests. To examine whether each variable significantly changed during Pre-SA, the present post-hoc tests examined the difference between the first subperiod（i.e., 1yr1Sem）and the subperiod just before SA（i.e., 3yr1Sem）, while the test of the difference between the other sets of subperiods was conducted only when it was necessary. The significance level was set at p=0.01 to protect against an inflated type 1 error.

3. Results

3-1. The speed and composite measures

Figure 1 shows the articulation rate（AR; the speed measure）and the speech rate（SR; the composite measure）as a function of the subperiods. It is shown that AR substantially increased from 1yr1Sem to 1yrSprV, decreased toward 2yrSumV, and

increased again to the level of 1yr2Sem, and then increased substantially between 2yrSpV to SA. The overall change was significant, $H(9)=32.5$, $p<0.001$, $\eta^2=0.118$[4]. The difference in AR was marginally significant between 1yr1Sem and 3yr1Sem, $U=98$, $z=-1.86$, $p=0.062$, $r=0.29$[5], while it was significant between 3yr1Sem and SA, $U=99$, $z=-2.70$, $p=0.007$, $r=0.41$. The results indicated that, overall, the speed of speaking increased to some degree during Pre-SA, while it significantly increased during SA.

It is also shown that SR followed a similar pattern as was shown in AR. The overall change was significant, $H(9)=58.1$, $p<0.001$, $\eta^2=0.246$. The difference was significant both between 1yr1Sem and 3yr1Sem, $U=71$, $z=-2.68$, $p=0.007$, $r=0.42$, and between 3yr1Sem and SA, $U=104$, $z=-2.57$, $p=0.01$, $r=0.39$. The difference between 1yrSumV and 1yr2Sem was marginally significant, $U=120$, $z=-2.48$, $p=0.013$, $r=0.37$. The results indicated that P1 became able to speak with a greater speed and fewer pauses both in Pre-SA and SA. It should be noted that the increase in SR was largely attributed to that between 1yrSumV and 1yr2Sem, and between 2yrSprV and SA. This suggested that the improvement was due to the training in the 2nd semester of the 1st year, and the 1st semester of the 3rd year. It should also be noted that the level of achievement at SA was still far below the NS means for both AR (267.5) and SR (227.7).

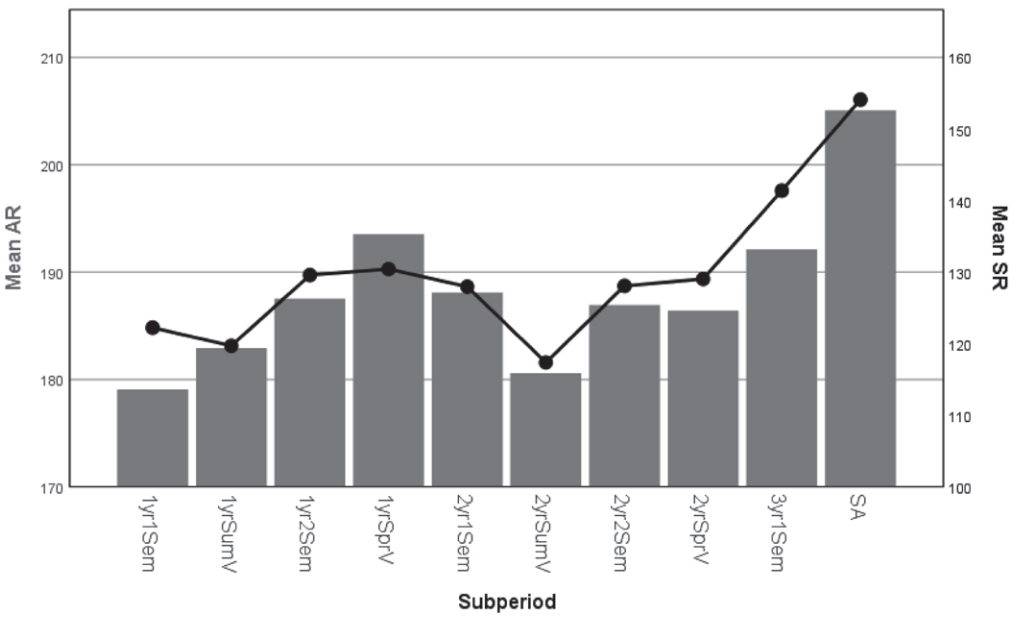

Figure 1: Mean (M) of the speed and composite measures averaged over narratives recorded within each sub-period in P1. AR=articulation rate; SR=speech rate. The bars indicate the mean AR, while the line the mean SR.

Table 1 shows the mean length of runs (MLoR; the composite measure) as a function of the subperiods with the average of native English speakers (NS). It followed the developmental pattern similar to SR. The overall change was significant, $H(9) = 50.9$, $p<0.001$, $\eta^2=0.210$. The difference was significant between 1yr1Sem and 3yr1Sem, $U=42$, $z = -3.56$, $p < .001$, $r=0.56$, while it was marginally significant between 3yr1Sem and SA, $U=109$, $z=-2.44$, $p=0.015$, $r=0.37$. The difference between 1yrSumV and 1yr2Sem was marginally significant, $U=130$, $z=-2.24$, $p=0.025$, $r=0.33$. The results indicated that, overall, P1 became able to produce a longer stretch of syllables between pauses in Pre-SA and SA. As was shown in SR, the substantial improvement was observed in the 2nd semester of the 1st year, the 1st semester of the 3rd year, and during SA.

Table 1: Mean (*M*) of the composite measure, the mean length of runs, averaged over narratives recorded within each sub-period in P1. MLoR=mean length of runs. NS=native speakers.

Subperiod		1yr1Sem	1yrSumV	1yr2Sem	1yrSprV	2yr1Sem	2yrSumV	2yr2Sem	2yrSprV	3yr1Sem	SA	NS
MLoR	*M*	4.5	4.7	5.2	5.1	4.9	4.6	4.9	5.0	5.7	6.5	14.9
	SD	0.8	0.7	0.7	0.7	0.7	0.6	0.7	0.6	0.8	1.0	4.7
	N	11	14	32	20	26	16	31	17	29	14	15

3-2. The breakdown measures

Table 2 shows the breakdown measures of the pause-phonation ratio (PauseRat) and the pause frequencies and durations related to the between-clause pauses as a function of the subperiods. First, PauseRat fluctuated between 30% and 35% until 2yrSprV and dropped up to SA. The overall change was significant, $H(9)=74.7$, $p<0.001$, $\eta^2=0.329$. The difference was significant between 1yr1Sem and 3yr1Sem, $U=70$, $z=-2.44$, $p=0.007$, $r=0.43$, but was not significant between 3yr1Sem and SA ($p>0.05$). The results indicated that the proportion of pauses in P1's narrative production largely decreased during the 1st semester of the 3rd year and continued to drop during SA.

The frequency of the between-clause pauses (PauseFreqBC) dropped gradually up to 2yrSumV, fluctuated until 2yrSprV, dropped sharply at 3yr1Sem, but only slightly decreased at SA. The overall decrease was significant, $H(9) = 33.1$, $p<0.001$, $\eta^2=0.121$, and so was the difference between 1yr1Sem and 3yr1Sem, $U=63$, $z=-2.94$, $p=0.003$, $r=0.46$. The duration of the between-clause pauses (PauseDruBC) fluctuated around 700ms and 1000ms and did not show a clear developmental pattern. The results indicated that P1 made significantly fewer between-clause pauses in Pre-SA.

The between-clause pauses were broken down into the clause-final pauses

Table 2: Mean (*M*) of the breakdown measures, averaged over narratives recorded within each sub-period in P1. PauseRat=pause duration ratio; PauseFreqBC=frequency of between-clause Pauses; PauseDurBC=Duration of between-clause pauses; PauseFreqCF=frequency of clause-final pauses; PauseDurCF=frequency of clause-final pauses; PauseFreqCI; frequency of clause-initial pauses; PauseDurCI=duration of clause-initial pauses; NS=native speakers.

Subperiod		1yr1Sem	1yrSumV	1yr2Sem	1yrSprV	2yr1Sem	2yrSumV	2yr2Sem	2yrSprV	3yr1Sem	SA	NS
PauseRat	*M*	32.0	35.0	31.0	33.0	32.0	35.0	32.0	31.0	27.0	25.0	15.0
	SD	6.0	5.0	3.0	4.0	3.0	4.0	3.0	4.0	3.0	3.0	5.0
PauseFreqBC	*M*	14.0	14.3	13.2	12.9	12.4	12.1	12.7	12.4	10.8	10.6	5.3
	SD	3.2	2.1	2.5	2.1	3.0	2.3	2.2	2.1	2.1	2.5	1.9
PauseDurBC	*M*	0.75	0.93	0.86	0.87	0.84	0.97	0.79	0.82	0.71	0.68	0.57
	SD	0.15	0.15	0.12	0.11	0.08	0.15	0.10	0.09	0.08	0.07	0.19
PauseFreqCF	*M*	7.1	7.8	7.2	6.2	6.3	6.4	7.0	6.8	6.0	6.0	2.4
	SD	1.7	1.8	1.7	1.5	1.2	1.1	1.4	1.3	1.5	1.3	1.0
PauseDurCF	*M*	0.69	1.10	1.00	1.05	0.98	1.08	0.88	0.89	0.77	0.75	0.62
	SD	0.12	0.24	0.17	0.13	0.13	0.17	0.17	0.10	0.09	0.08	0.22
PauseFreqCI	*M*	6.9	6.5	6.0	6.7	6.1	5.8	5.7	5.6	4.8	4.6	3.0
	SD	2.3	2.0	2.3	1.8	2.6	2.0	2.1	1.6	1.6	2.0	1.5
PauseDurCI	*M*	0.80	0.73	0.69	0.68	0.70	0.81	0.68	0.75	0.63	0.58	0.52
	SD	0.24	0.23	0.16	0.12	0.13	0.20	0.15	0.18	0.14	0.08	0.22
	N	11	14	32	20	26	16	31	17	29	14	15

(PauseFre/DurCF) and clause-initial pauses (PauseFre/DurCI), as described above. As shown in Table 1, PauseFreqCF showed a gradual decline toward SA, but the difference was not significant either between 1yr1Sem and 3yr1Sem or between 3yr1Sem and SA ($p>0.05$). PauseDurCF did not show a clear pattern of development. In contrast, PauseFreqCI declined toward SA, especially in the latter half of the overall period. The overall decline was marginally significant, $H(9)=20.5$, $p=0.015$, $\eta^2=0.058$, while the difference between 1yr1Sem and 3yr1Sem was significant, $U=75$, $z=-2.56$, $p=0.001$, $r=0.40$. PauseDurCI also showed a decline, with some fluctuations, toward SA. The overall change was significant, $H(9)=23.3$, $p=0.006$, $\eta^2=0.072$. The difference between1yr1Sem and 3yr1Sem was marginally significant, $U=90$, $z=-2.11$, $p=0.035$, $r=0.33$, but was not significant between 3yr1Sem and SA ($p>0.05$). The results showed that the clause-initial pauses showed a clearer developmental decline than the clause-final pauses both in terms of frequency and duration.

Table 3 shows the pause frequencies and durations associated with the within-clause pauses as a function of the subperiods. First of all, the frequency of the within-clause

pauses (PauseFreqWC) dropped sharply from 1yr1Sem to 1yr2Sem, rebounded toward 2yrSumV, and decreased again toward SA. Overall, however, it showed a downward developmental trend. The overall change was significant, $H(9)=44.6$, $p<0.001$, $\eta^2=0.178$. The difference was significant between both 1yr1Sem and 3yr1Sem, $U=64$, $z=-2.89$, $p=0.004$, $r=0.46$, and between 3yr1Sem and SA, $U=102$, $z=-2.62$, $p=0.009$, $r=0.40$. In addition, the difference was significant between 1yr1Sem and 1yr2Sem, $U=45$, $z=-3.65$, $p<0.001$, $r=0.56$. The duration of the within-clause pauses (PauseDurWC) fluctuated between 570ms and 700ms through the course of the subperiods. It should be noted that that the observed range of durations was not substantially higher than that of the NS mean (i.e., 520ms). The results indicated that the within-clause pauses significantly decreased in frequency both in Pre-SA and SA. It was also indicated that the frequency significantly decreased by the end of the 1st year of the program.

Table 3: Mean (*M*) of the breakdown measures, averaged over narratives recorded within each sub-period in P1. PauseFreqWC=frequency of within-clause Pauses; PauseDurWC= duration of within-clause pauses; PauseFreqWCPB=frequency of within-clause, between-phrase pauses; PauseDurWCPB=duration of within-clause, between-phrase pauses PauseFreqWCWPB=frequency of within-clause, within-phrase pauses; PauseDurWCWPB; duration of within-clause, within-phrase pauses; NS=native speakers.

Subperiod		1yr1Sem	1yrSumV	1yr2Sem	1yrSprV	2yr1Sem	2yrSumV	2yr2Sem	2yrSprV	3yr1Sem	SA	NS
PauseFreqWC	*M*	9.2	6.3	5.6	6.2	7.6	9.1	7.6	7.1	6.4	4.6	2.1
	SD	2.5	3.0	1.9	2.5	2.4	2.9	3.2	2.5	2.5	1.4	1.4
PauseDurWC	*M*	0.60	0.69	0.57	0.64	0.62	0.70	0.66	0.60	0.57	0.58	0.52
	SD	0.10	0.24	0.13	0.13	0.10	0.15	0.14	0.10	0.10	0.08	0.13
PauseFreqWCPB	*M*	7.4	5.4	4.8	5.5	6.3	7.9	5.8	5.7	5.5	3.6	1.5
	SD	2.5	2.9	1.8	2.3	2.1	2.9	2.8	2.6	2.2	1.4	1.0
PauseDurWCPB	*M*	0.60	0.75	0.58	0.67	0.62	0.75	0.68	0.63	0.59	0.60	0.59
	SD	0.12	0.25	0.15	0.13	0.12	0.17	0.19	0.13	0.11	0.10	0.16
PauseFreqWCWPB	*M*	1.8	0.9	0.8	0.7	1.3	1.2	1.8	1.5	1.0	1.0	0.6
	SD	1.6	1.0	0.6	0.9	1.1	0.9	1.1	1.2	1.1	1.0	0.6
PauseDurWCWPB	*M*	0.64	0.54	0.64	0.46	0.62	0.52	0.65	0.57	0.53	0.58	0.54
	SD	0.25	0.13	0.31	0.20	0.17	0.17	0.19	0.14	0.20	0.15	0.18
	N	11	14	32	20	26	16	31	17	29	14	15

The within-clause pauses were broken down into the within-clause, between-phrase pauses (PauseFreq/DurWCPB) and the within-clause, within-phrase pauses (PauseFreq/DurWCWPB). As shown in Table 3, a comparison of the frequency of the two types of pauses showed that the proportion of the between-phrase pauses was much higher than

that of the within-phrase pauses. It showed that a larger proportion of the within-clause pauses were located between the phrase boundaries (e.g., I lived // in the house) rather than within a phrase (e.g., I lived in the // house). PauseFreqWCPB dropped sharply toward 1yr2Sem, increased toward 2yrSumV, decreased again toward 3yr1Sem, and dropped sharply at SA. The overall change was significant, $H(9)=33.6$, $p<0.001$, $\eta^2=0.123$. The difference between 1yr1Sem and 3yr1Sem was marginally significant, $U=91$, $z=-2.09$, $p=0.037$, $r=0.33$, while the difference between 3yr1Sem and SA was significant, $U=100$, $z=-2.68$, $p=0.007$, $r=0.41$. PauseDurWCPB did not show a developmental trend, fluctuating between 580ms and 750ms, which was not substantially different from that of the NS mean (i.e., 590ms). The overall results found that the frequency of the within-clause, between-phrase pauses showed a strong tendency to decrease in Pre-SA and significantly decreased in SA. PauseFreqWCWPB showed a pattern similar to PauseFreqWCPB, but the difference was not significant either between 1yr1Sem and 3yr1Sem, or 3yr1Sem and SA ($p>0.05$). The observed range of PauseDurWCPB was not substantially different from the NS mean (i.e., 540ms). The results showed that the frequency of the within-clause, within-phrase pauses showed only a tendency to decrease toward the NS mean in Pre-SA, and that a further decline was not observed in SA.

3-3. The repair measures and fillers

Table 4 shows the frequency of the repair measures and of fillers. The overall frequency of repairs (RepairFreq) sharply dropped between 1yr1Sem and 1yrSumV, where the frequency was lower than the NS mean (i.e., 1.8). The difference between 1yr1Sem and 3yr1sem was significant, $U=64$, $z=-2.89$, $p=0.004$, $r=0.46$, indicating that, overall, the frequency of repairs significantly decreased in Pre-SA. The frequency of repairs was broken into that of false starts (FSFreq), repeats (RPFreq), and self-corrections (SCFreq). Although the overall frequency is relatively small, FSFreq increased in 3yr1Sem and SA, indicating that P1 rephrased the sentence structures more frequently than the previous subperiods. RPFreq dropped sharply between 1yr1Sem and 1yrSumV, indicating that P1 was able to substantially decrease the repeats during 1yr1sem. The difference between 1yr1Sem and 3yr1sem was significant, $U=63$, $z=-3.20$, $p=0.001$, $r=0.51$. SCFreq was relatively small, and it did not show a clear developmental pattern. Finally, the frequency of fillers (FLFreq) dropped sharply between 1yr1stSem and 1yrSumV, indicating that P1's use of fillers decreased during 1yr1Sem. In fact, FLFreq at 1yr1Sem was lower than the NS mean (i.e., 3.9), indicating that the NS used

Table 4: Mean (*M*) of the repair measures, averaged over narratives recorded within each subperiod in P1. RepairFreq=frequency of repairs; FSFreq=frequency of false starts; RPFreq=frequency of repeats; SCFreq=frequency of self-corrections; FLFreq=frequency of fillers; NS=native speakers.

Subperiod		1yr1Sem	1yrSumV	1yr2Sem	1yrSprV	2yr1Sem	2yrSumV	2yr2Sem	2yrSprV	3yr1Sem	SA	NS
RepairFreq	*M*	2.7	0.3	0.5	1.3	1.5	1.5	1.3	1.2	1.3	1.6	1.8
	SD	1.4	0.4	0.6	1.2	1.1	1.5	0.9	0.7	0.8	1.2	1.7
FSFreq	*M*	0.4	0.0	0.0	0.0	0.1	0.1	0.4	0.3	0.5	0.6	0.1
	SD	0.6	0.0	0.0	0.2	0.3	0.2	0.6	0.4	0.6	0.6	0.2
RPFreq	*M*	1.4	0.2	0.3	0.6	0.2	0.7	0.2	0.4	0.3	0.2	1.2
	SD	1.3	0.4	0.5	0.7	0.4	1.0	0.4	0.5	0.4	0.3	1.4
SCFreq	*M*	0.9	0.1	0.3	0.7	1.2	0.7	0.8	0.6	0.6	0.8	0.6
	SD	1.0	0.3	0.4	0.7	1.0	0.9	0.9	0.6	0.6	1.0	0.6
FLFreq	*M*	3.0	0.1	0.0	0.1	0.1	0.3	0.2	0.1	0.2	0.0	3.9
	SD	0.7	0.3	0.1	0.4	0.2	0.8	0.4	0.2	0.4	0.2	2.4
	N	11	14	32	20	26	16	31	17	29	14	15

fillers more frequently than P1 on the average.

4. Discussion and Conclusion

4-1. The speed and composite measures

The present article was designed to examine how the fluency measures (i.e., the speed, composite, breakdown, repair fluency) changed in the FI settings before SA (i.e., Pre-SA) and during about a month under the SA settings (i.e., SA), and how different fluency measures interacted with each other. It was found that all the fluency measures significantly improved from the beginning of Pre-SA through the first part of SA. The results clearly indicated that the participant (P1) became able to speak with a greater speed, a larger chunk of syllables/words, fewer between- and within-clause pauses, and fewer repairs. The overall changes of fluency may best be represented by the speech rate (SR; the composite measure), as it is a combined measure of the speed and breakdown fluency.

As shown in Figure 1, the fluency of P1 improved between the 1st and the 2nd semester (although the statistical significance between 1yrSumV and 1yr2Sem was marginal: p=0.013; r=0.37, the effect size is considered to be larger than medium). A closer examination of the data showed that P1's SR increased from 116.4 to 131.6 from the 1st to the 9th month (i.e., from April to January) when P1 took the formal English classes.

This suggested that the improvement in fluency was partly due to the speaking training called "Timed Pair Practice" employed in the GCP program's speaking class (cf. Pipe & Tsushima, 2021b, 2022a, 2022b, 2022 (in press)). The fluency, however, remained at the same level through 2yr1Sem. This apparent plateau was followed by a noticeable decrease in fluency in 2yrSumV. A closer examination of the data suggested that the decline in her performance was not due to relatively more difficult topics of the narratives, relatively higher levels of vocabulary, or more complex grammatical structures used in the narratives. According to an interview with P1 conducted by the author, this deterioration in her performance might be explained by a motivational factor coupled with a limited exposure to spoken English. She said that she had lost her motivation to study English hard during the summer vacation because SA, which was supposed to take place in the 2nd semester of the 2nd year, was postponed because of the COVID-19. A similar account might be applied to P1's significant increase in SR at 3yr1Sem as it had been decided that the students would be able to go to SA just before the beginning of 3yr1Sem. The results underscore the importance of the students' motivation to improve their speaking skills to prepare for SA.

It was found that the articulation rate (AR; the speed measure) changed almost in tandem with SR (see Figure 1). It was also found that the mean length of runs (MLoR; the composite measure) closely approximated the changes shown by SR (see Table 1). In terms of the production models described above, (Kormos, 2006), the composite and speed measures are associated with all the stages of the production process (i.e., conceptualization, formulation, and articulation stage). The results suggested that, in Pre-SA and SA, P1 was able to speed up the process of producing a preverbal plan in the conceptualizer, that of lexical access/retrieval and syntactic/morpho-phonological encoding in the formulator, and that of executing articulatory movements. The present findings were commensurate with the results of the previous case studies in which the longitudinal development of fluency was examined in the FI settings (Tsushima, 2019, 2020, 2021). They were also compatible with the findings of a recent study (Pipe & Tsushima, 2022a) which examined the group data (N=11) on the development of fluency during the first year of GCP. In the study, all the speed and composite measures (i.e., AR, SR, MLoR) significantly improved from the 1st half of the 1st semester to the 2nd half of the 2nd semester.

4-2. Improvement in the breakdown measures

4-2-1. Pause-phonation rate

The present results found that pause-phonation rate (PauseRat) significantly decreased in Pre-SA (see Table 2). It was indicated that P1 was able to decrease the overall amount of the pause time relative to the speaking time through the speaking classes and speech training. The significant decrease in PauseRat was due to the decrease in frequency and duration of both the between-clause pauses and within-clause pauses to be discussed just below.

4-2-2. Between-clause pauses

It was found that the frequency of the between-clause pauses (PauseFreqBC) significantly decreased in Pre-SA (see Table 2). When the between-clause pauses were divided into the clause-final pauses (PauseFreq/DurCF) and clause-initial pauses (PauseFreq/DurCI), the significant decrease in the frequency of PauseFreqBC was largely due to that of the latter (PauseFreqCI). This suggested that the number of the sentences did not change substantially throughout the subperiods. Rather, it was the number of pauses which occurred before the start of the following sentence that significantly decreased. Consistent with the data on frequency, it was also found that the duration of the clause-initial pauses (PauseDurCI) significantly decreased in Pre-SA, although that of the clause-final pauses (PauseDurCF) did not. Previous research has suggested that the between-clause pauses are related with activation of background knowledge, planning of content, and ordering of the conveyed information (Butterworth, 1975; Götz, 2013; Lambert et al., 2017). According to the speech production model described above, they are primarily associated with the conceptualizer (Kormos, 2006). Presumably, the decrease in the frequency and duration of the clause-initial pauses might be due to the increasingly efficient process of producing the preverbal plan.

4-2-3. Within-clause pauses

The present results found that the frequency of the within-clause pauses (PauseFreqWC) significantly decreased in Pre-SA (see Table 3). As noted above, PauseDurWC was already close to the NS mean at the beginning of Pre-SA. The findings suggested that P1 became able to produce a longer chunk of phrases without a pause through the speaking classes and speech training. The present analyses divided the within-clause pauses into the within-clause, phrasal pauses and the within-clause, within-

phrasal pauses. It was found that the proportion of the frequency of the within-clause, phrasal pauses (PauseFreqWCPB) was much higher than that of the within-clause, within-phrasal pauses (PauseFreqWCWPB) (see Table 3). PauseFreqWCPB significantly decreased from 1yr1Sem to SA, indicating that they developed the ability not to make pauses at the phrasal boundaries within a clause (e.g., before a noun, adjectival, or prepositional phrase). On the other hand, the frequency of occurrence for PauseFreqWCWPB was relatively low even in 1yr1Sem. In fact, P1's PauseFreqWCWPB in 1yr1Sem was 1.8 times per 100 syllables, indicating that she made pauses within a noun, adjectival, or prepositional phrase only occasionally. However, PauseFreqWCWPB showed a tendency to decrease toward the NS mean, suggesting that P1 reduced the probability of making such pauses in Pre-SA. The previous research has suggested that the within-clause pauses are primarily associated with the formulator stage in L2 speech production (Götz, 2013; Kormos, 2006; Skehan et al., 2012). The present results suggested that the process of formulating a phrase structure by means of lexical access/retrieval and syntactic/morpho-phonological encoding became more efficient and speedier in Pre-SA.

4-2-4. Between-clause and within-clause pauses

As shown just above, both the between-clause and within-clause pauses significantly decreased in Pre-SA. The present results, however, showed that the within-clause pauses declined earlier than the between-clause pauses. A further examination of the data revealed the following. The difference in the frequency of the clause-initial pauses (PauseFreqCI) between 1yr1Sem and 1yr2Sem was not significant ($p>0.05$), while the difference in the within-clause pauses (PauseFreqWC) between the same subperiods was significant, $U=45$, $z=-3.65$, $p<0.001$, $r=0.56$ (see Table 2 and 3). The result indicated that the frequency of the within-clause pauses decreased at a greater rate than that of the between-clause pauses during the 1st year of the program. This finding suggested that, at an earlier stage of L2 speech production learning, P1 improved the ability to reduce the number of pauses located at phrase boundaries within a clause. The present finding is compatible with that of the previous study (Tsushima, 2021) where a similar relation between the between- and within-clause pauses in terms of the rate of development was shown by a participant with a less advanced speaking ability (i.e., PauseFreqWC=23.9 in the initial subperiod of three months). This may support the hypothesis that the decrease in the frequency of the within-clause pauses may begin earlier than that of the between-

clause pauses across L2 learners of different proficiency levels.

4-3. The repair measures

It was found that the overall frequency of repairs significantly changed (p<0.001, see Table 4). The frequency was relatively low (i.e., 2.7 times per 100 syllables) at 1yr1Sem. In the following subperiod (i.e., 1yrSemV), it substantially decreased to 0.3, which was lower than the NS mean (i.e., 1.8). Then, it increased and fluctuated around 1.5. When the overall frequency was broken down into three categories (i.e., false starts (FSFreq), Repeats (RPFreq), and self-corrections (SCFreq), RPFreq was the highest in P1 at 1yr1Sem (i.e., 1.4). However, it decreased to below once in 100 syllables in the following subperiods, indicating that P1 rarely used this type of repairs.

The present results, combined with those of a previous study (Tsushima, 2021), suggested that there was individual variability in L2 learners' use of repairs in terms of the overall frequency, the developmental pattern, and the types of repairs. As a result, the frequency of repairs may not be a good indicator of the learner's proficiency (Pipe & Tsushima, 2022a; Saito et al., 2018; Tavakoli et al., 2020; Tsushima, 2021) as exemplified by the data on P1. Even among the native speakers, FepairFreq varied from 0.0 to 5.6 per 100 syllables, as the relatively high *SD* indicated (see Table 4). In the previous study which used a similar case study procedure (Tsushima, 2021), one participant with lower proficiency rarely used repairs after a speech training period of one year (i.e., SR=105.9; RPFreq=0.24 in the last subperiod), while the other participant with higher proficiency used repairs after a speech training period of one and half years (i.e., SR=144.3; RPFreq=3.57 in the last subperiod). The individual differences might be partly due to idiosyncratic speaking styles. For example, one learner may prepare the structure of the following phrase or sentence before starting to speak by using a relatively long pause, while another leaner may just utter a subject, and prepare the structure of the following verb phrase by repeating the subject. The individual difference might also be due to the ability to self-monitor different aspects of the speaker's speech. Less advanced learners might have limited resources to self-monitor, for example, grammatical mistakes, which results in fewer self-corrections in the middle of a sentence. One other hand, more advanced learners might notice online a variety of aspects of their speech including the syntactic structure, grammatical errors, mispronunciation of segments, misplacement of stress. This might lead to an increase in the frequency of false starts and/or self-corrections.

4-4. Filled pauses

It was found that the frequency of filled pauses (FLFreq) significantly changed across the subperiods (p<0.001, see Table 4). FLFreq was 3.0 in 1yr1Sem, which was lower than the NS mean (i.e., 3.9), and decreased to 0.1 in the next subperiod (i.e., 1yrSemV) and fluctuated between 0.1 and 0.3 in the following subperiods. This finding indicated that P1 rarely used a filled pause except for the first subperiod. The data on the native speakers showed that FLFreq varied from 0.82 to 9.76, as indicated by the relatively high *SD*. Some learners may use them to "fill the time" while constructing the following phrase or sentence, while others may just use a silent pause for the same purpose. In the group data of the learners in GCP, FLFreq decreased from 12.6 to 6.1 from the first quarter to the last quarter during the first year (Pipe & Tsushima, 2022a), although the difference was not statistically significant due to relatively high individual variability. In the production model (Kormos, 2006), the decrease in FLFreq might be due to the improvement in the speed and efficiency of processing in both the conceptualizer and the formulator. As learners improve the ability to formulate a preverbal plan in the conceptualizer, a phrase structure through lexical access/retrieval and syntactic/morpho-phonological encoding processes and a phonetic plan through phonetic and phonological encoding processes in the formulator, they may become able not to use the fillers before uttering a sentence or a phrase.

4-5. Teaching implications

First, the present findings imply that L2 leaners fluency can be significantly improved in the FI settings if some effective teaching technique is employed (Pipe & Tsushima, 2021a, 2021b, 2022a, 2022b, 2022 (in press); Tsushima, 2020, 2021). During the first year of GCP, Timed-Pair-Practice Framework (TPPF) was employed, which improved a number of the fluency measures in P1 to varying degrees. DeKeyser (2007) argues that, to improve L2 fluency, learners should be provided with the opportunity to automatize syntactic, morphological, and phonetic encoding processes and the use of prefabricated language units (i.e., formulaic language). To facilitate the automatization processes and the use of formulaic language, they should practice repeating a set of syntactic and morphological rules (i.e., procedural knowledge of the encoding processes) under demanding conditions where they express real ideas and intentions. These conditions appear to be satisfied in the teaching procedure of TPPF (see Pipe & Tsushima, 2021b for a description of the procedure).

Second, the present findings imply that motivational factors played a prominent role in the improvement of fluency (cf. Saito, Dewaele, & Hanzawa, 2017). In P1, SR stayed at the same level during 2yr1Sem, decreased substantially at 2yrSemV, returned to the prior level in 2yr2sem, remained at the same level until 2yrSprSem. Apart from the drop in 2yrSumV, the plateau in terms of the fluency level continued for as long as one year. SR and the other fluency measures substantially improved in 3yr1Sem after it was decided that the group was able to go to the study abroad and continued to improve during SA. As described above, the drop in fluency in 2yrSumV appears to be largely due to lack of her motivation to study English. In contrast, the sharp increase in fluency appears to be closely related with the improvement of her motivation to develop English speaking skills in anticipation for going to Australia. The increase in motivation in 3yr1Sem appears to be primarily intrinsic (Gardner & Lambert, 1972) in that she had the spontaneous tendency "to seek out novelty and challenges, to extend and exercise one's capacity, to explore, and to learn" (Ryan & Deci, 2000, p. 70). Dörnyei (2002) states, "… motivation is never static but is constantly increasing or decreasing, depending on the various social influences surrounding action, the learner's appraisal of these influences and the action control operations the learner carries out on such motivational content (p.156)." There appears to be little doubt that the changes in P1's motivation in Pre-SA were strongly influenced by her future expectations in terms of speaking English in SA, and her urge to improve the fluency to prepare for the study abroad.

Finally, the present findings, together with the previous studies, indicated that it takes a great amount of time for L2 learners to improve the fluency in a spontaneous speech task such as the narrative production task employed in the present study. As shown above, P1 experienced a year of plateau with a drop in performance in the middle. The findings imply that foreign language instructors, as well as L2 learners, should be prepared for the long span of a learning period to significantly improve their fluency. It is recommended that instructors understand an existence of a relatively long period of plateau where little improvement is observed, and encourage the learners to continue practicing until the following period of substantial improvement takes place in fluency.

4-6. Limitations of the study

First, as is the case with other case studies, the results may not be generalized to other individuals or groups who differ in language proficiencies, learning environments, learning purposes and some other factors. Currently, data from a group of students in

GCP are being obtained from the beginning of the 1st year through a period after SA. The ongoing study is expected to provide further data on the effects of English study in Pre-SA and SA among English learners with varying levels of proficiency. Second, the present study does not include the data on perceived fluency, that is, how the fluency of the produced narratives is evaluated by native speakers of English. Such data would provide important information on the relation between the improvement in the utterance fluency and that of the perceived fluency (cf. Saito et al., 2018). Future research on this point is certainly required. Finally, the present data are based on the narrative production task which is repeated a great number of times. It might be the case that, as the participants became skilled exclusively in this type of task, and the data might not represent their ability in other spontaneous speech situations. To guard against this possibility, other types of spontaneous tasks (e.g., an impromptu speech task) have been conducted before SA and will be conducted after SA. The data are expected to supplement the current data.

4-7. Concluding remarks

The present article is an interim report of a case study on the development of fluency during the preparation period before SA and a beginning part of SA in one participant in the study abroad program (i.e., GCP). The present data amply demonstrated that the fluency significantly improved in production of English in a spontaneous narrative production task before SA. Regarding the effects of SA, the current data at the time of writing (i.e., approximately two months and a half after SA began) have shown that the fluency has been further improving. The study is going to examine not just the data during SA, but also whether and how the improved fluency is going to be retained at least three months after SA. It is hoped that, combined with the ongoing study which examines the group data, the present study is going to provide valuable information about the way various aspects of L2 fluency improve in the FI and SA settings. It is also hoped that the case study will reveal important information about individual differences that the averaged group data may not be able to capture.

Acknowledgement

The present research was supported by a research grant (21-23) from Tokyo Keizai University in the academic year of 2021. The author greatly appreciates the work of Jason Pipe in collecting the data from native speakers of English.

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）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.
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.

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.

Butterworth, B.（1975）. Hesitation and semantic planning in speech, *Journal of Psycholinguistic Research*（4）, 75–87.

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

Dörnyei, Z.（2002）. The motivational basis of language learning tasks. In P. Robinson（Ed.）, *Individual Differences and Instructed Language Learning*（pp. 137–158）. Amsterdam: John Benjamins.

Freed, B.（1995）. What makes us think that students who study abroad become fluent? In B. Freed（Ed.）, *Second Language Acquisition in a Study Abroad.*（pp. 123–148）. Amsterdam: John Benjamins.

Gardner, R. C., & Lambert, W. E.（1972）. *Attitudes and Motivation in Second Language Learning.* Rowley, MA: Newbury House Publishers.

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.

Götz, S.（2013）. *Fluency in Native and Nonnative English Speech*（Vol. 53）. Philadelphia/Amsterdam: John Benjamins.

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.（2021）. Task preparation and second language speech production. *TESOL Quarterly, 55*（2）, 331–365. doi: https://doi.org/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.

Lennon, P.（1990）. Investigating fluency in EFL: A quantitative approach. *Language Learning, 40*（3）, 387–417.

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, March). *Improved fluency through the Timed-pair-practice framework.* Paper presented at the Asian Conference on Language 2021, Tokyo.

Pipe, J., & Tsushima, T. (2022a). Improving proficiency in fluency and pronunciation in a Japanese context. *Korea TESOL Journal, 18* (1), 143–177.

Pipe, J., & Tsushima, T. (2022b). Progress in fluency and pronunciation through the Timed-Pair-Practice Framework. *Journal of Humanities & Natural Sciences, 150*, 115–165.

Pipe, J., & Tsushima, T. (2022 (in press)). Progress in Fluency through the Timed-Pair-Practice Framework. *Journal of Humanities & Natural Sciences* (151).

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*, 68–78. doi:10.1037/0003-066X.55.1.68

Saito, K., Dewaele, J.-M., & Hanzawa, K. (2017). A Longitudinal Investigation of the Relationship between Motivation and Late Second Language Speech Learning in Classroom Settings. *Language and Speech, 60* (4), 614–632. doi:10.1177/0023830916687793

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.

Skehan, P., Xiaoyue, B., Qian, L., & Wang, Z. (2012). The task is not enough: Processing approaches to task-based performance. *Language Teaching Research, 16* (2), 170–187. doi:10.1177/1362168811428414

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

Towell, R., Hawkins, R., & Bazergui, N. (1996). The development of fluency in advanced learners of French. *Applied Linguistics, 17* (1), 84–119.

Tsushima, T. (2008). How are nonnative perception and production ability related during perception and production training? *Journal of the Japan Society for Speech Sciences, 9*, 29–47.

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.

Tsushima, T. (2021). A case study: longitudinal development of oral fluency in production of monologic narratives among two adult Japanese learners of English. *Journal of Humanities & Natural Sciences* (149), 69–92.

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.