Progress in Fluency through the Timed-Pair-Practice Framework

Jason Pipe and Teruaki Tsushima[1)]

ABSTRACT

This is the third research paper to determine how Timed-Pair-Practice (TPP) can be a catalyst to improving fluency in speech. Previous research had shown clear improvement in this area in terms of speed, pausing and repair and natural alteration in the pause location to between-clause boundaries to reflect a more native-like speech production. Through the inclusion of TPP to a new cohort of low-intermediate Japanese students (G1: n=11), this paper hoped to confirm whether the previous results of improved fluency from another research group (G2: n=12) could be replicated. Data was collected from both groups, along with a control group (CG: n=17) and a native group (NG: n=15). By using a triad of composite measures to determine fluency, it became evident that TPP had a positive influence on speech production of participants, clearly outperforming the control group but not to the extent of the native group. This would indicate that despite cultural linguistic concerns and inappropriate pedagogic methodologies introduced at high school, students were able to participate fully in their paired discussions with their peers. Furthermore, G1 could outperform G2 in their progress in their fluency over the academic year which may have resulted from fine-tuning the application of TPP in classes.

Keywords: timed-pair-practice, fluency, parallel processing, cognitive processing

1. Introduction

As part of the Global Career Program, first university students at a Japanese university have been receiving classes to prepare them for studying and working abroad for a six-month period. By focusing on strengthening their communicative skills through

the introduction of framework, Timed-Pair-Practice (TPP) (refer to Pipe & Tsushima, 2020, for further explanation of TPP), it was hoped that students could improve their intelligibility and comprehensibility when expressing themselves in English while living abroad (Ahangari et al., 2015; Derwing et al. 2012; Koike 2014; Robinson et al. 2012; Vicsi & Szaszák, 2010; Yenkimaleki & van Heuven, 2016). Although speech production is generally evaluated by the descriptors of grammar, vocabulary, fluency, and pronunciation, the former two descriptors were not explicitly taught in class, unless necessary, as these were aspects already covered in their high school education through the grammar translation method (Kikuchi & Browne, 2009). It was, therefore, assumed that students had a "reasonable" grasp of English grammar and vocabulary. However, fluency and pronunciation were areas that required attention. Previous research on TPP has had a positive influence on fluency with improvements in the speed of language production, reduction in pausing and less reliance in the use of repairs such as filled pauses and repeated phrases as proficiency improved (Pipe & Tsushima, 2021/2022). This study, therefore, aimed to replicate similar positive changes in regards to fluency through quantitative analysis.

2. Challenges to Fluency

Conversing in a language fluently not only requires being able to provide a large vocabulary but speaking error-free and with native-like pronunciation (Bosker et al., 2013). Such spoken language should also be produced with less hesitation and relative ease, and at an adequate speed (Tavakoli et al., 2020). Finally, fluency is noted for its flow, continuity, automaticity and smoothness (Koponen & Riggenbach, 2000). However, to become more fluent, the speaker has to consider a wide range of factors which can be bewildering to the non-native speaker. The issues stem from a variety of unresolved issues.

First, students are compounded by incredulous prioritization of direct translation methods of vocabulary and grammar and at high school. This is to prepare the students for their university entrance examinations (Butler, 2015; Løfsgaard, 2015; Steele & Zhang, 2016, Tahira, 2012). Teachers, therefore, have unsurprisingly concentrated on the teaching and memorization of aspects of the English language to pass these exams. Unfortunately, such attention on these examinations has resulted in little consideration to the methodology of natural second language acquisition nor the application of these aspects in

real conversation in task-based learning. Furthermore, with Japanese as the main medium of instruction (Steele & Zhang, 2016; Takahashi, 2005), and a shortage of qualified school teachers licensed to teach communicative English at a secondary-school level (Japan Times, 2019; Nakata, 2011), students have naturally struggled in their English education. As a consequence, despite a 6-year minimum period of English study at regular school, it is understandable why the English education in Japan is ranked low. For the past decade, Japanese students of English have been categorized as being in the "low proficiency" band (Education First, 2020) and has resulted in this country being placed 55th out of one hundred countries and one of the lowest among Asian countries (Educational Testing Service, 2019). The above factors have, therefore, severely impeded the level of involvement students invest in their conversational classes.

Another factor that has adversely affected students' ability to communicate effectively in English is as a direct result of their culture. Cultural linguistic concerns that affect communication classes include efforts in maintaining collective communication system (Hofstede et al., 2010), hierarchical respect for their teacher and peers (Banks, 2016), and harmonious relationships within the class (Nisbett & Masuda 2003). Other cultural factors include avoiding shame in making mistakes in their second language abilities in front of others (Kawamura et al., 2006); and discouragement in conveying individual opinions (Ting-Toomey & Chung, 2005). Such cultural concerns have profoundly impacted the level of engagement by Japanese students to converse with their peers in their English communicative classes.

Finally, the sheer level of processing and encoding linguistic information when conversing in a second language puts heavy demands on working memory resources. Following the comprehensive model of Levelt's (1989) four-stage speech model of language processing and production (de Bot, 1992; Doe, 2017; Kormos, 2006; Pulvermuller, 2002; Segalowitz, 2010; Skehan, 2009), it is expected that L2 learners in general are less automatic in accessing their declarative knowledge of syntactic, lexical, and phonological rules than L1 speakers (Kormos, 2006). This is particularly so for the Japanese students as a result of issues with culture and teaching methodology as mentioned in the previous paragraph. Dysfluencies will understandably result as L2 learners resort to slowing down of speech, pausing, or using filled pauses to maintain conversation (Tavakoli, 2011) due to slow conscious *serial processing*. This is as a result of gaps in linguistic knowledge, interference from L1 transfer, and a lack of automaticity (Tavakoli & Wright, 2019; Pipe & Tsushima, 2021). When considering the amount of language processing expected in the

classroom, it is expected that L2 learners are less automatic in accessing their declarative knowledge of syntactic, lexical and phonological rules (Kormos, 2006) i.e. inefficient, effortful grammatical, lexical and phonological encoding (Mora & Levkina, 2017). In fact, this bottleneck form of *controlled processing* in language production (Ellis, 2005) is further limited by the individual's short-term memory capacity (Shiffrin & Schneider, 1977). This would certainly add light to why Japanese students have performed poorly in their spoken English as this stage demands much working memory resources (Hao & Othman, 2021). Unfortunately, few strategies have been used to relieve this cognitive burden by effectively improving the L2 learner's knowledge of the target language or resource deficit (Dornyei & Scott, 1997, Kormos, 2006). When considering inappropriate pedagogic methodologies introduced at high school and cultural linguistic concerns, it is with good reason that students struggle with processing and encoding linguistic information in L2 communication.

3. Encouraging Fluency with TPP

With effective preparation, however, students should benefit from TPP and the robust positive effects of repetition of tasks (Ahmadian & Tavakoli, 2011; Lambert et al., 2017; Sheppard & Ellis, 2018; Wang, 2014). In time, one should notice improvement in cognitive processing (Derwing, et al., 2008; Segalowitz, 2003; Segalowitz & Freed, 2004), and faster reaction time in the testing (e.g., Ammar, 2008; Lyster & Izquierdo, 2009) as students develop the notion of *parallel processing* (Kormos, 2006, Lambert et al., 2021, Skehan 2014). In other words, through the repetitive processing of tasks (Lambert et al., 2017) or recursive conversations (Brown, 2014; Kindt & Bowyer, 2018), students will become more able not only in their competency in their spoken English (Bowyer, 2019) but also more focused on working on two stages of speech production more simultaneously as one aspect of production, such as the conceptualization and formulation stages or greater automation of encoding processes. Through conscious effort, L2 learners could overcome such problems in communication by spending much attention in carrying out and practicing such processes consciously (Kormos, 2006).

However, applying the technique of *parallel processing* will take time as L2 learners' language knowledge/competence is rarely complete, and students will struggle with limits on working memory resources (Hao & Othman, 2021). Learning must consider a comfortable balance between encouraging active cognitive processing through meaningful

negotiation so as to allow incoming information to be interpreted, reorganized and integrated with formerly acquired knowledge (Fiorella & Mayer, 2015); while pushing students to practice and absorb this information through sufficient repetitive practice and thereby result in easing the pressure on the learner's limited working memory in the long run (Ellis, 2005). As a result, the repetitious nature of TPP provides a workable trade-off between the two. However, students will find relevant aspects of schematic and content memory have been activated and remembered from the first task as a result of constantly performing repetitious speaking tasks. This will lead to greater efficiency in L2 language processing and allow students to subsequently direct more cognitive resources for allocating attention to formulating linguistic forms accurately (Fukuta, 2016). This will also compensate for the learner's weakness in cognitive aptitude (Skehan, 2016) and enable students to apply more parallel-like processing to tasks by practicing encoding of utterances in a more real time manner (Lambert et al., 2021), which will lead to improved mental lexicon organisation and lexical retrieval through more enriched conceptual specifications (de Groot, 1995; Kormos, 2006), increasing proficiency of the declarative knowledge of L2 rules (Kormos, 2006), and progress in their overall L2 proficiency (Sample & Michel, 2014). Such *automatic processing* (Ellis, 2005) would, therefore, lead to the short-term memory capacity being virtually unaffected (Shiffrin & Schneider, 1977) as speech production would require "significantly less efforts or attention," and "relatively immune to disruption" (Segalowitz, 2003: 382). Such improvement would be noted by improved fluency overall (Ahmadian & Tavakoli, 2011; de Jong & Perfetti, 2011; Lambert et al., 2017; Thai & Boers, 2016). However, a shift from strenuous serial processing to more efficient parallel processing can be potentially captured by observing less frequent pausing, especially between clausal boundaries (BCB), due to improved conceptualization (Lambert et al. 2017) and other dysfluencies resulting from repair due to improved control over L2 knowledge (Kormos, 2006; Lambert et al. 2017; Pipe &Tsushima, 2021; Saito et al., 2018; Skehan, 2016; Skehan & Shum, 2017; Tavakoli & Wright, 2019).

4. Research focus

To determine the effectiveness of the TPP framework, this study compared and contrasted the present research group (G1) with the previous research (G2) (conducted by Pipe & Tsushima, 2021) as well as a control group (CG) and native group (NG) by focusing on the following specific research question:

How did the speed, breakdown and repair measures show student progress in their fluency of English?

5. Methodology and Methods

5. 1 Participants

The participants (G1: n=11; G2: n=11) were first year students from a private university in Tokyo. Despite having a minimum of six years of learning, their TOEIC scores varied from 400 to 755 while Versant scores ranged from 29 to 48. This would indicate CEFR levels of the experimental group being between lower B1 and upper A1 which would mean their English ability can be categorized as high beginner to intermediate. Their data was contrasted with a control group of Japanese students (CG: n=17) who attended a general English communication class which did not include instruction using TPP and a native group of English speakers (NG: n=15).

5. 2 Timed-Pair-Practice Procedure

For motivational purposes (Porter, 1999), the students were required to prepare a 250-word response and 20 questions on a topic chosen by themselves. These students were then expected to practice asking these questions in pairs. After changing partners in subsequent rounds during this practice stage, the students became able to ask more appropriate questions and maintain longer conversations. After sufficient practice, students were then evaluated in the testing stage in which two students, picked at random, would be asked to provide another conversation on the same topic chosen. Through these practice and testing rounds, it was hoped that students would develop greater fluency by spending less processing time on the formulation, articulation, and self-monitoring stages of these aspects of the spoken language.

5. 3 Data Elicitation

Students performed a weekly narrative production task but only 12 out of the 24 recordings were chosen in order to save time in data analysis. This task consisted of a one-minute spontaneous monologue explaining what happened in each student's week. Dialogue recordings were not considered as there were concerns in regards to overlap and unclaimed pauses between turns. There would also appear to be little difference between the performance in the monologues and dialogues in terms of frequency and

location, speed and length of pauses (Tavakoli, 2016). For the purposes of simplicity and reliability, therefore, it was considered prudent to analyse individual narratives to measure each student's spontaneous speaking ability.

All student utterances were recorded at a resolution of 16 bits with a sampling rate of 44.1 Hz by a PCM recorder through a high-quality microphone placed approximately 20cm from the mouth of the speaker. This data was transferred to a computer in which the recorded sounds were low-pass filtered at 8,000 Hz, normalized, and analyzed by sound analysis software, *Praat* (Boersma & Weenink, 2014).

5. 4　Notation Analysis

To determine the location of the pause at sentence level, extracted data was analysed through the notational form of the syntactic AS-unit (Analysis of Speech Unit) as this would seem the most effective way to codify spoken data (Moser, 2010) due to its flexibility (Foster et al., 2000) and simplicity (Ellis & Barkhuizen. 2005). AS-unit refers to a single speaker's utterance consisting of an independent clause or sub-clausal unit together with any subordinate clause (s) (Foster et al., 2000). Once this spoken data had been codified through the AS-unit, pausing could be evaluated in terms of frequency and mean length measures at the clause level to establish pausing at BCB and NCB.

5. 5　Analysis Procedure

To determine the fluency of the research groups, only the recordings taken from the odd weeks were analysed due to expediency. To establish a base-rate for the control group, the recordings of the fluency were made twice in the 5^{th} and 11^{th} week for both terms. All recordings were transferred onto a digital format, in which the second author transcribed a sampled one-minute speech and matched each lexical item to the recording on the software, *Praat*. Then, the acoustic data were segmented into consonants, vowels, and pauses, and duration of each portion was measured.

5 .6　Triad Fluency Measures

Qualitative data has been collected by investigating utterance fluency (Housen & Kuiken, 2009; Tavakoli & Skehan, 2005; Segalowitz, 2016). This would relate to the acoustically measurable aspects of fluency in uttered speech by examining speed, pausing and repair (Appendix 1). Speed had been evaluated by four key measures: Speech Rate (SR), Articulation Rate (AR), Phonation-time Ratio (PhonRat) and Length of Runs

(MLoR), to provide more credible results (Tsushima, 2018; Valls-Ferrer & Mora, 2014). Pausing was measured by the Pause Rate (PR) to determine how speakers at lower proficiency levels rely on longer silent pauses to process and produce speech (Tavakoli et al., 2020), and pause location within non-clausal boundaries (NCB) and between-clausal boundaries (BCB). Repair had been analysed by examining filled pauses (FP) which looks at set phases or sounds to maintain some output; false starts (FS) in which an utterance is attempted but either abandoned altogether or reformulated in some way (Foster et al., 2000); repeats (RR) where the speaker repeats previously produced speech (Maclay & Osgood, 1959); and self-correction (SC) when the speaker identifies an error either during or immediately following production and stops and reformulates the speech (Levelt, 1989). This paper applied these fluency measures (Tavakoli et al., 2020) to determine how fluency changed over the academic year. These measures should illuminate the underlying thought processes involved when producing utterances (Huensch & Tracy–Ventura, 2017; Hunter, 2017; Tavakoli & Hunter, 2018) and unfold the complex nature of fluency (Bosker et al., 2013; de Jong et al., 2012; Kahng, 2014; Kormos, 2006; Pipe & Tsushima, 2021; Skehan, 2015; Tavakoli et al., 2020).

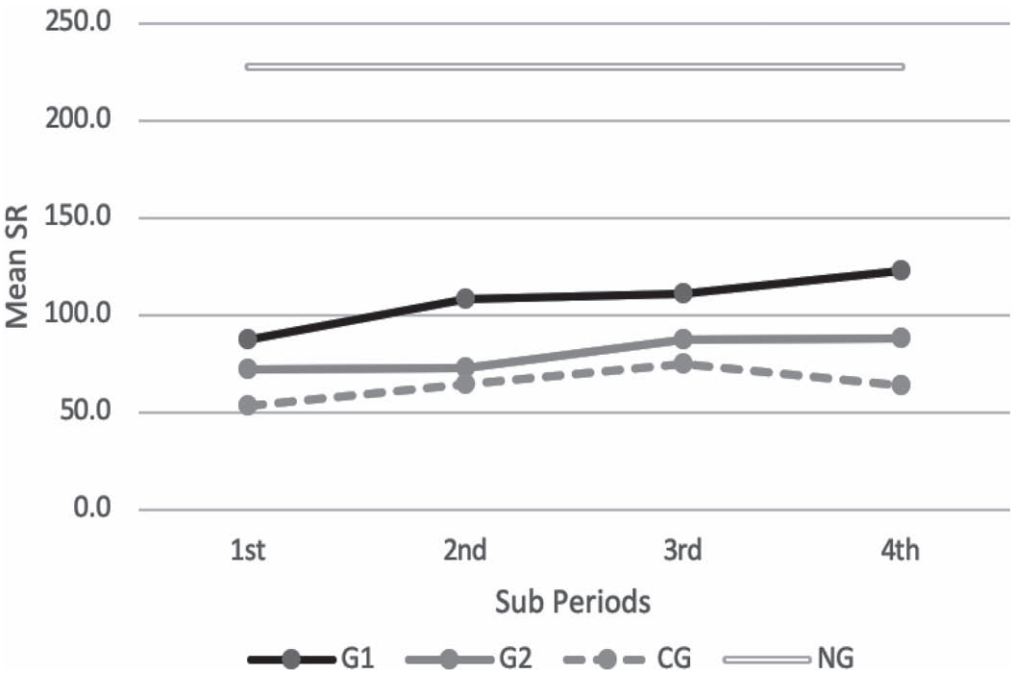
5. 7 Statistical Analysis Procedure

The fluency measures analysed twelve data points in the research groups but were later averaged into four quarters in order to compare the results with the four data points in the control group and the one data point for the native group. To test the statistical significance of these data sets, a repeated-measures ANOVA was run on each measure separately. When the assumption of sphericity was not met, Greenhouse-Geisser correction was used. The statistical analyses of the control group showed that all the fluency measures, including speed, composite, breakdown, and repairs measures, were not significant, $p>0.05$, except for SR, $p=0.05$, $\eta^2=0.22$, indicating that the group improved very little during the semester. Therefore, the average data will be presented in the result section.

6. 1. Speed Performance

Looking at the speed of language output, the research group showed modest improvement in their speed in speech rate (SR), articulation rate (AR), mean length of run (MLoR) and phonation-time ratio (PhonRat) compared to the control group but there is still further progress required to match the level of a native speaker.

6. 1. 1　Speech Rate

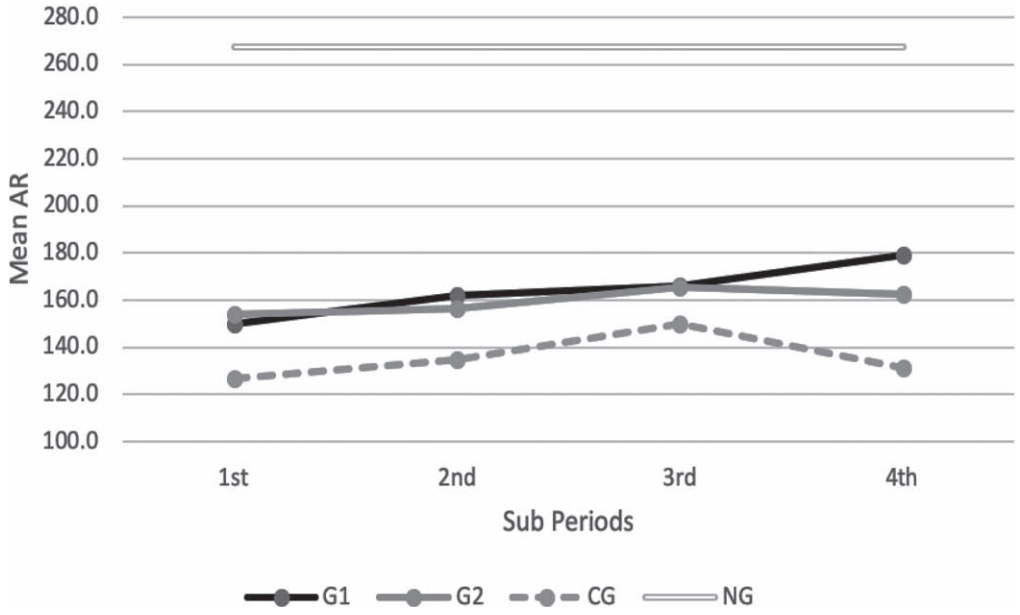


Graph 1: Speech rate performance of research, control and native groups

For G1, SR significantly increased over the academic year, F (5, 60)=18.8, p<0.001, η^2=0.94, increasing constantly from 87.2 syllables/min in the first quarter to 122.4 syllables/min in the final quarter (an increase of 40.4%) which is a marked improvement when contrasted with the control group's mean average of 64.0 syllables/min. Furthermore, this is a significant improvement from G2 which also improved from 72.0 to 88.8 syllables/min (an increase of 21.9%). However, the native group spoke at a much faster rate of 227.7 syllables/min.

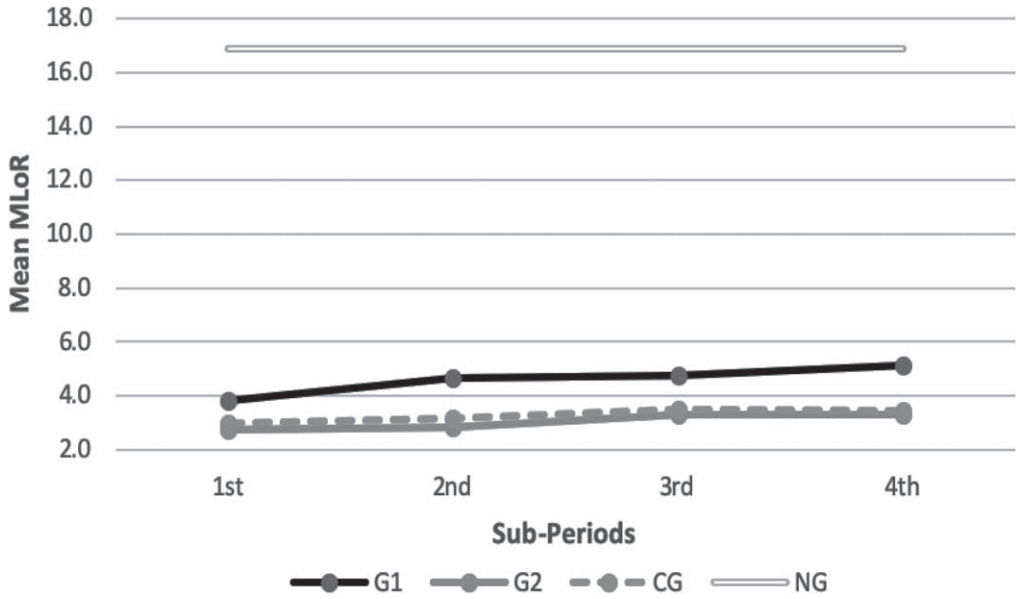
6. 1. 2　Articulation Rate

A pattern was also observed in AR as G1 significantly increased their rate, F (5, 60) =4.7, p=0.001, η^2=0.28, from 149.8 syllables/min to 179.1 syllables/min (an increase of 19.5%). Again, when compared to the results of the control group's mean average of 137.1 syllables/min, there is clear progress. Again, this is a significant improvement from G2 which only improved from 153.8 to 162.3 syllables/min (an increase of 5.5%). However, despite such gains, the native group spoke at a much faster rate of 267.5 syllables/min.



Graph 2: Articulation rate performance of research, control and native groups

6. 1. 3 Mean length of run

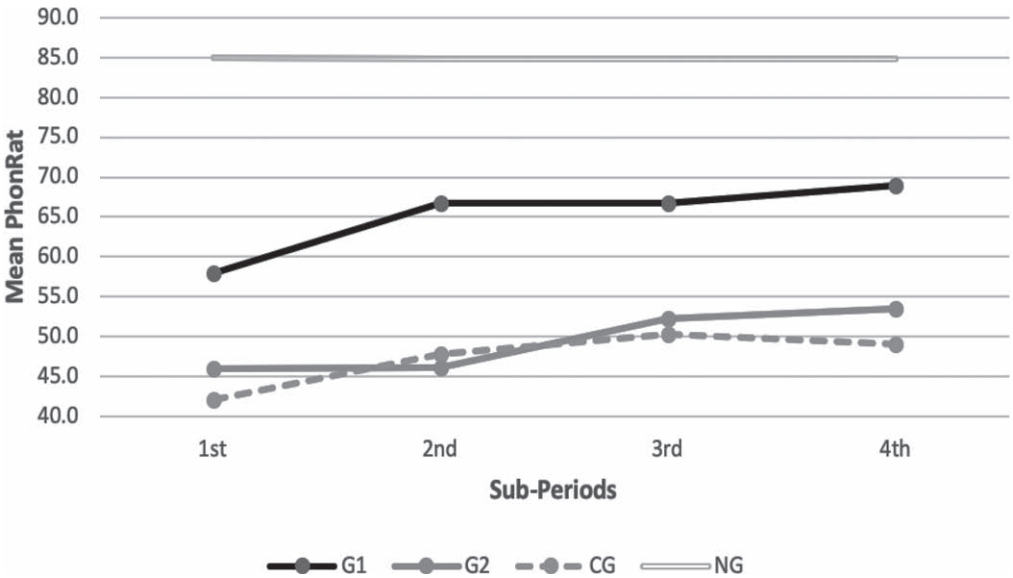


Graph 3: Mean length of run of research, control and native groups

*MLoR units were measured in syllables/run for G1, CG and NG while G2 was measured in words/run

Focusing on the third speed measure, the MLoR, the research group provided lengthier runs with a significant increase, F (2.9, 35.2)=4.6, p=0.009, η^2=0.28, from 3.8 to 5.1 syllables/run (an increase of 34.2%) which outperformed the control group which managed an average of 3.2 syllables/run. G2 was measured in words/minute. It is noted that there is a disparity in the units of measurement as words contain one or more syllables so it is expected that the figures will be lower for G2. However, this group also showed significant improvement from 2.8 to 3.3 words/run (an increase of 17.9%). Despite such improvement by both G1 and G2, a clear gap still remains regarding the ability to maintain lengthier utterances with native speakers averaging a run of 16.9 syllables/run.

6. 1. 4　Phonation-Time Ratio



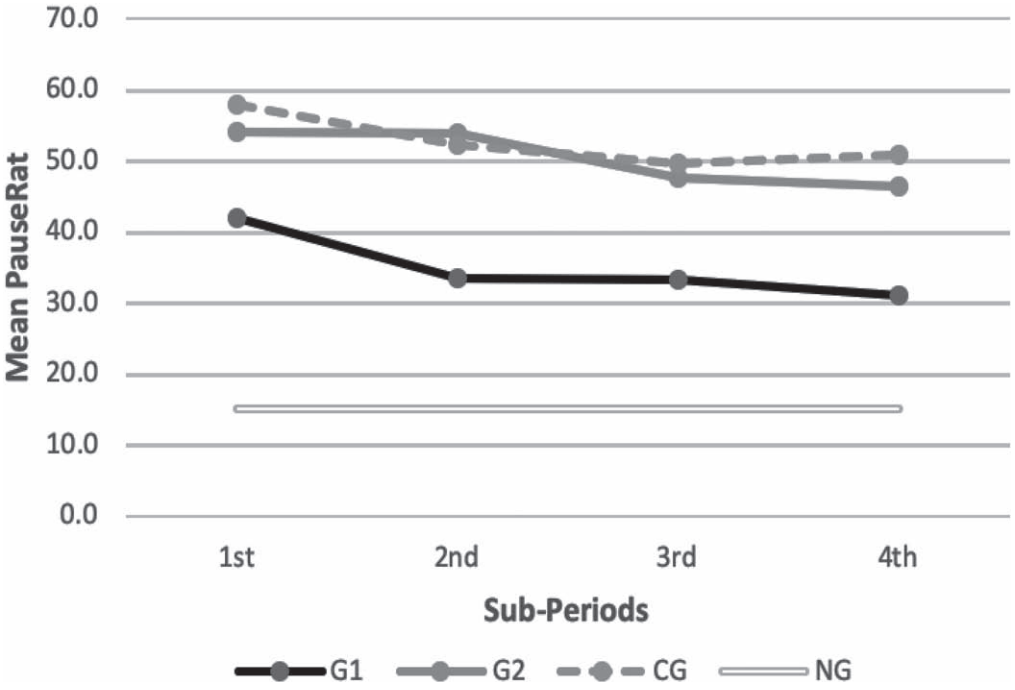
Graph 4: Phonation-time ratio of research, control and native groups

The final speed measure looks as the percentage of speech production and, again, there is a significant improvement, F (5, 60)=13.0, p<0.001, η^2=0.52. G1's PhonRat began the first quarter at 58.0% and increased to 69.0% by the end of the semester (increase of 19.0%). The control group, on the other hand, showed a considerably lower average PhonRat percentage of 46.7%. There is also a significant improvement from G2 which only improved from 46.0% to 53.5% (an increase of 16.3%). However, despite the research group edging closer, the PhonRat of the natives was much higher at 85.0%.

6. 2 Pause Performance

Drawing attention towards pausing when attempting to produce language output, G1 showed modest improvement with a reduction in pausing overall, especially when compared to the control group. There was a drop in: pause rate (PR), pausing at both the non-clausal boundaries (NCB), and the between-clausal boundary (BCB). However, there is still further progress required to match the level of a native speaker.

6. 2. 1 Pause Rate



Graph 5: Pause rate ratio of research, control and native groups

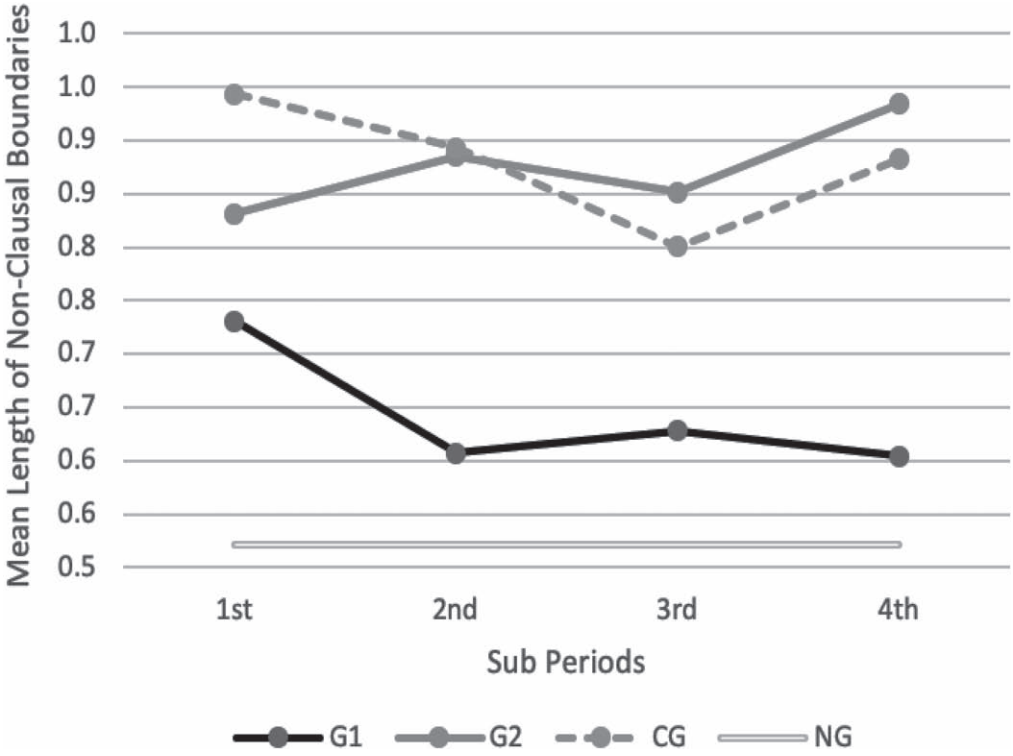
If we draw our attention to the amount of pausing by each group, we can see the reverse of PhonRat and, again, there is a significant improvement, F (5, 60)=13.0, p<0.001, η^2=0.52. The amount of pausing at the first quarter was little less than half the time taken to speak at 41.9% but this mean average significantly fell to 31.1% by the final week (a decrease of 25.8%). The control group was pausing considerably more in the recorded data at a mean PauseRat of 53.3%. Again, G1 clearly outperformed G2 which reduced the rate of pausing from 54.0% to 46.5% (a decrease of 13.9%). However, despite progress made by the research group, the rate of pausing was much higher than the native group

of 15.0%.

6. 2. 2　Pause Location within the unit of spoken language

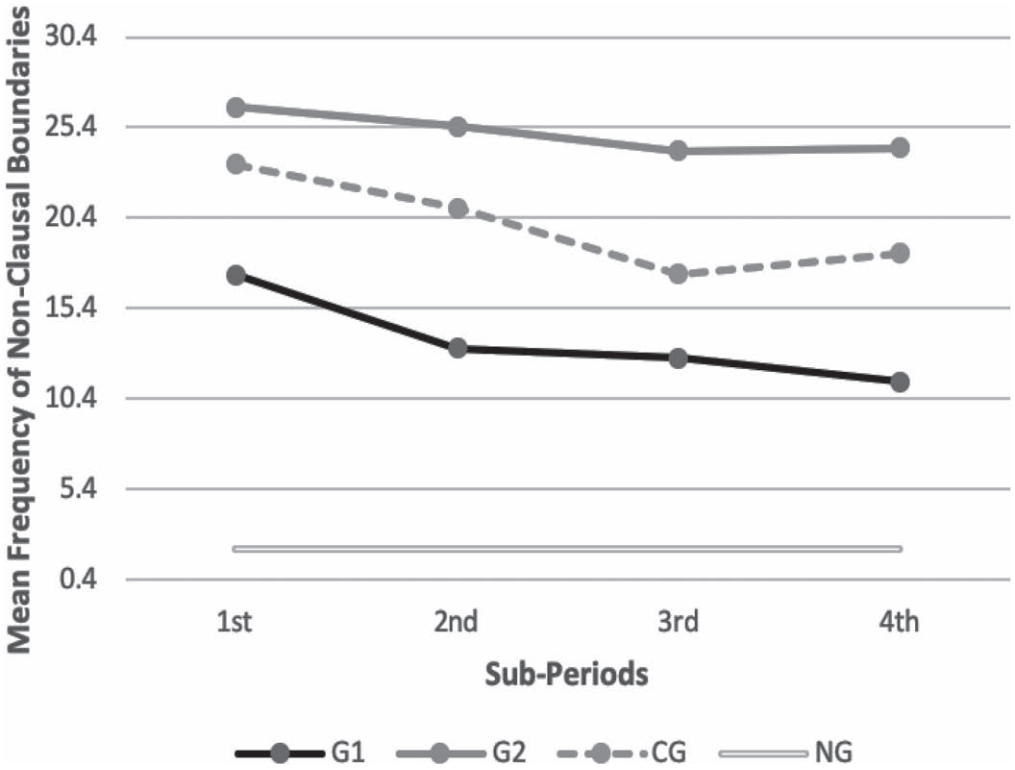
If we turn our attention to the pause location within the unit of spoken language, the following information will provide insight into pausing at the non-clausal boundary (NCB) and at between-clausal boundary (BCB).

6. 2. 2. 1　Pauses at Non-Clausal Boundaries



Graph 6: Mean of length of non-clausal boundary pausing of research, control and native groups

Looking at the mean length of pausing at the non-clausal boundaries (NCB), it would appear that G1 made overall progress with constant reduction in the mean length of NCB pausing, F (5, 60)=4.0, p=0.003, η^2=0.25, from 0.73 to 0.60 seconds by the final test (a decrease of 17.8%) which seemed to resemble closer to the native level of 0.52 seconds and a clear movement away from the control group with an average NCB pause duration of 0.88 seconds. G2, on the other hand, in fact increased the rate of pausing from 0.83 to 0.93 seconds (an increase of 12.0%).

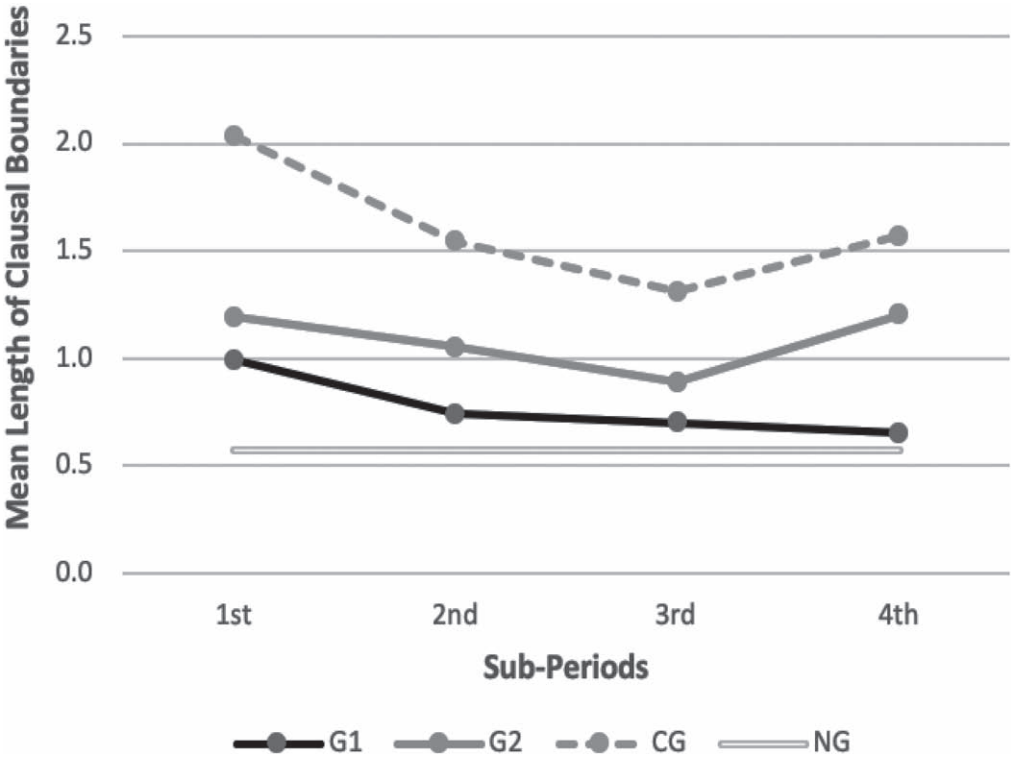


Graph 7: Mean of frequency of non-clausal boundary pausing/100 Syllables of research, control and native groups

Although similarly matching the initial results of the control group of 20.6 pauses/100 syllables, G1 significantly reduced the number of NCB pauses, F (3.0, 36.1)=6.8, p=0.001, η^2=0.36, from 17.2 pauses/100 syllables in the first quarter to 11.3 pauses/100 syllables by the final quarter (decrease of 34.3%). Again, G1 outperformed G2. G2 managed to reduce the rate of pausing slightly from 26.5 to 24.3 pauses/100 syllables (a decrease of 8.3%) which is higher than CG. However, despite the success of G1, the native group length of NCB pausing was considerably less at 2.1 seconds pauses/100 syllables.

6. 2. 2. 2 Pauses at Clausal Boundaries

The mean length of pause at between-clausal boundary (BCB) for the research groups significantly decreased, F (1.9, 22.0)=13.0, p<0.001, η^2=0.51. In the first quarter, the mean length was 0.99 seconds and 0.65 seconds by the final quarter (an overall decrease of 34.3%). This would indicate promising changes in length of pauses again as G1 not only performed better when compared to G2, which actually increased the rate of



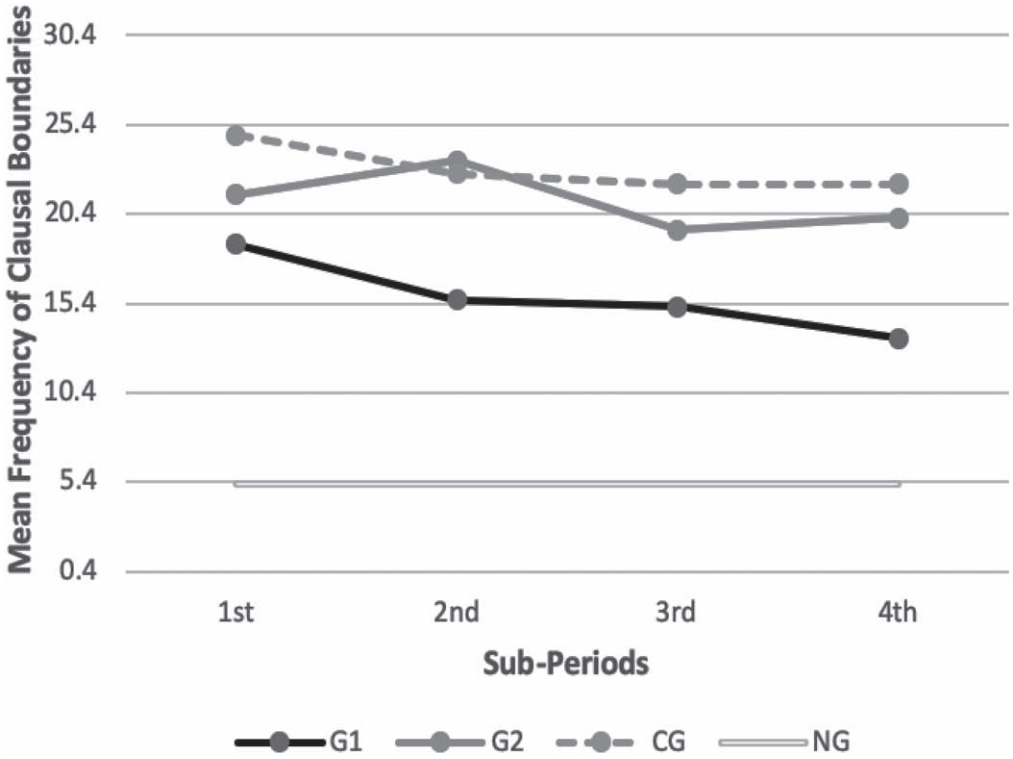
Graph 8: Mean of length of between-clausal boundary pausing of research, control and native groups

pausing overall from 1.19 to 1.21 seconds (an increase of 1.7%), but also reflected more closely with BCB pause length of the native level of 0.57 seconds. The control group, on the other hand, showed a lengthier average mean of 1.63 seconds.

Regarding the frequency of BCB pauses, again a similar pattern emerges. G1 showed progress in almost each test, starting at 18.7 pauses/100 syllables in the first quarter, falling to and 13.4 pauses/100 syllables by the final quarter (an overall decrease of 28.3%). Although the overall decline was marginally significant, F (2.6, 39.7)=2.7, p=0.07, η^2=0.18, and still not close to the native group which seldomly paused at BCB with a rate of 5.3 pauses/100 syllables, the results outperformed G2 which decreased the rate of pausing from 21.5 to 20.2 pauses/100 syllables (a decrease of 6.0%). The control group paused more often than the research group with an average rate of 23.2 pauses/100 syllables.

6. 3　Repair performance

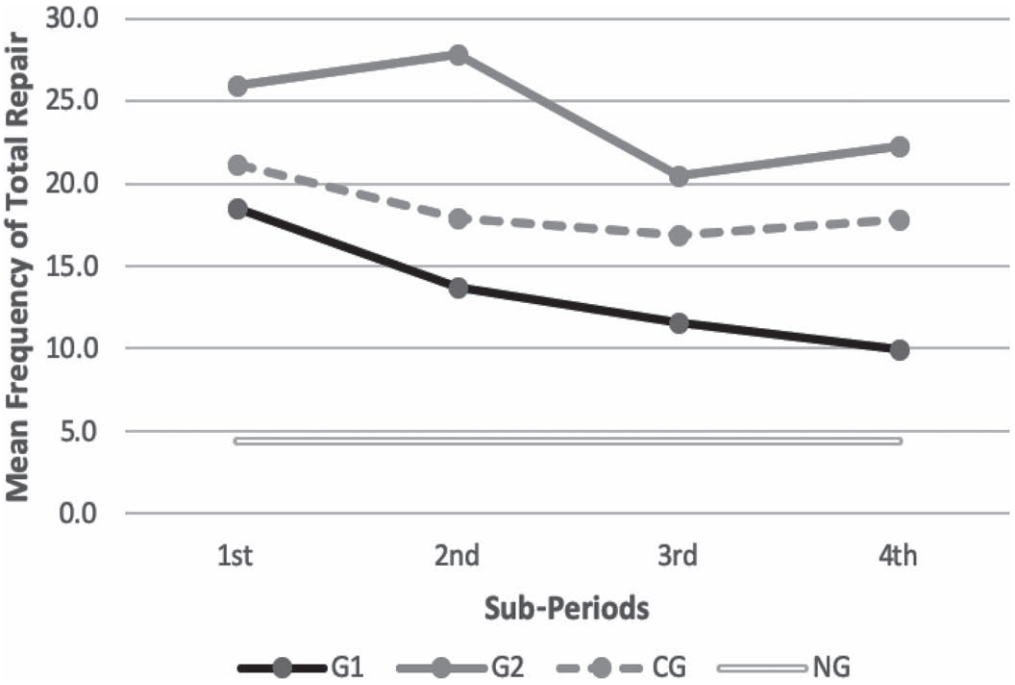
*Note that self-correction data was not included in the total repairs of G1, CG and NG as this was not



Graph 9: Mean of Frequency of between-clausal boundary pausing/100 syllables of research, control and native groups

measured in G2's data

Looking at the final fluency measure (graph 9), it would appear that G1 used this strategy consistently less when maintaining their utterances which would indicate greater fluency still. This group started at 18.6 repairs/100 syllables in the first quarter and 9.9 repairs/100 syllables by the final quarter (a decrease of 46.8%). G2 would appear to be either more ambitious or more challenged to maintain their utterances. The frequency of total repairs varied throughout the year – ranging from 20.5 repairs/100 syllables in the third quarter to 27.9 repairs/100 syllables in the second quarter. Surprisingly, CG would seem to have performed better than G2 with an average of 18.4 repairs/100 syllables but one must also take into account the speed and pausing measures. However, despite the progress made by G1, the native group seldomly had to reply on repair with only 4.3 repairs/100 syllables.



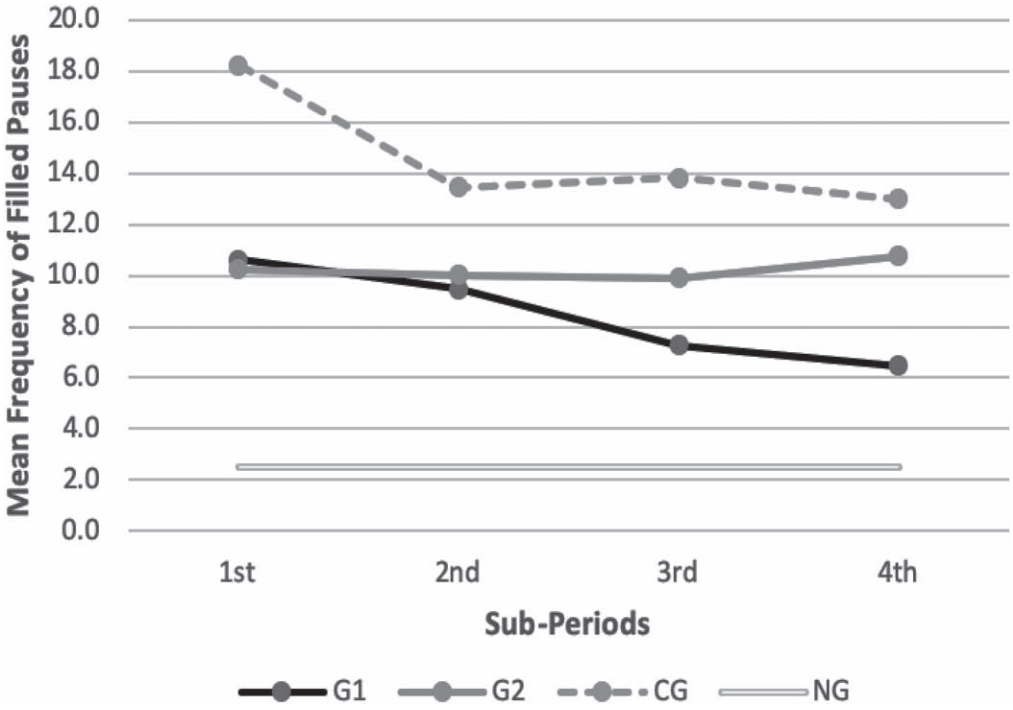
Graph 10: Mean frequency of total repairs/100 syllables of research, control and native groups

6. 3. 1　Filled Pauses

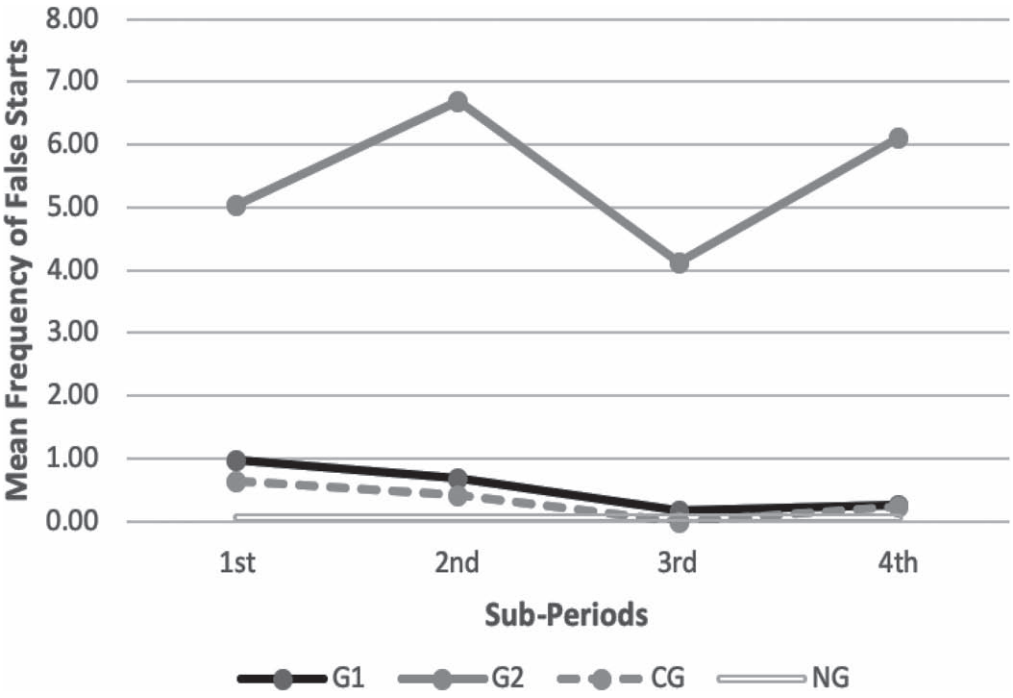
It is noted that G1 relied on filled pauses less over the first semester to maintain their utterances. In the initial quarter, this group averaged 10.6 filled pauses/100 syllables but this fell to 6.5 filled pauses/100 syllables by the end quarter (an overall decrease of 38.7%), although the decline was not statistically significant, F (2.2, 26.3)=2.19, p=0.128, η^2 =0.16. G2 would seem to use this repair strategy less, increasing slightly from 10.2 to 10.8 filled pauses/100 syllables in the final quarter (a net increase of 5.9%). The control group, on the other hand, managed an even higher rate with an average of 15.5 filled pauses/100 syllables. The native group, however, seldomly relied on this strategy to maintain their utterances with only 2.5 filled pauses/100 syllables.

6. 3. 2　False Starts

False starts would again demonstrate improvement in the G1's ability to maintain their utterances. Initially, G1 made an average of 0.98 false starts/100 syllables in the first quarter and this dropped to a minimum of 0.18 false starts/100 syllables in the third quarter but slightly increased to 0.27 false starts/100 syllables by the end of the course (a net decrease of 72.4%). Despite progress, the decrease was not statistically significant, F



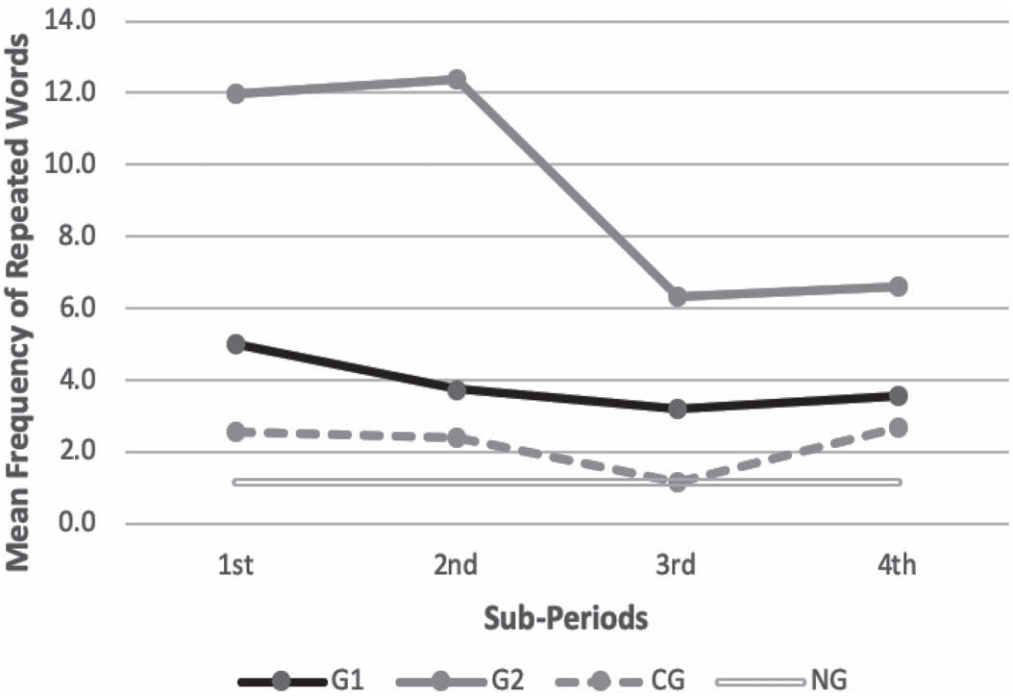
Graph 11: Mean frequency of filled pauses/100 syllables of research, control and native groups



Graph 12: Mean frequency of false starts/100 syllables of research, control and native groups

(5, 60)=1.7, p=0.139, η^2 =0.13. The control group seemed to rely on this strategy less with 0.5 false starts/100 syllables but not to the extent of the native group which used only 0.1 false starts/100 syllables. However, G2 again would seem to be more adventurous or struggled to maintain their utterance with much higher fluctuations in the use of this strategy with a low of 4.13 false starts/100 syllables in the third quarter to a high of 6.69 false starts/100 syllables in the second quarter (an overall increase of 21.4%).

6. 3. 3　Repeated Words

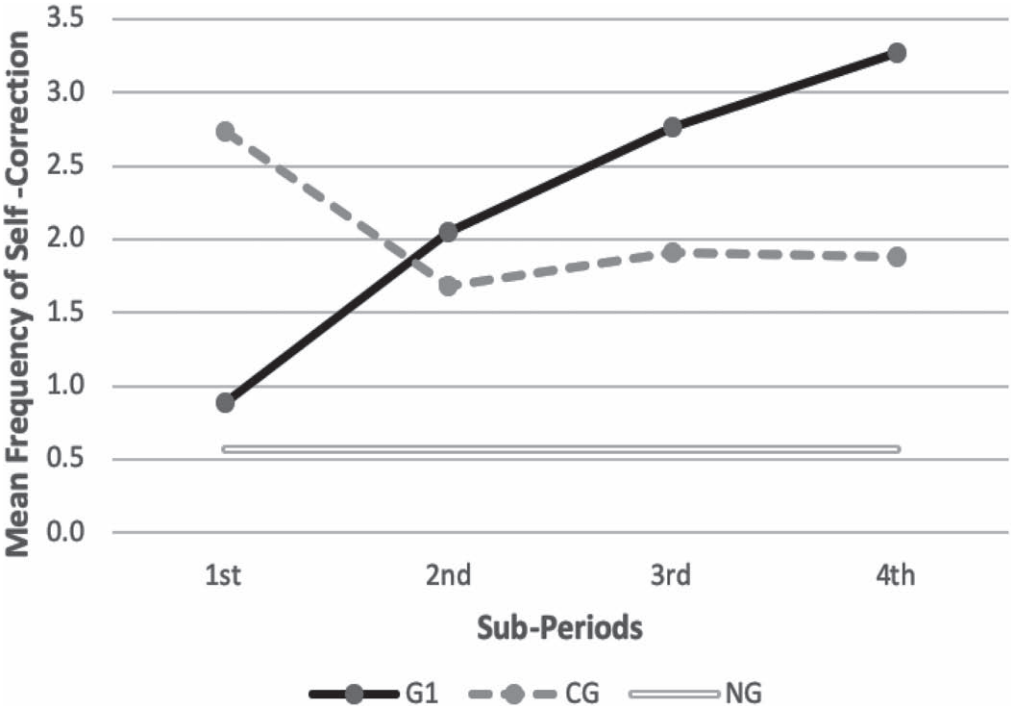


Graph 13: Repeats/100 syllables of research, control and native groups

Again, a similar pattern is observed by G1 as this group reduced their reliance on repeats to maintain their utterances. This group used 5.0 repeats/100 syllables in the first quarter and this fell to 3.2 in the third quarter, only to rise slightly to 3.6 repeats/100 syllables by the end of the semester (a decrease of 28.0%), F (2.3, 30)=10.3, p<0.001, η^2 =0.46. The control group, again, relied on this strategy less with an average of 2.0 repeats/100 syllables and the native group even lesser with 1.2 repeats/100 syllables. G2 also displayed a similar pattern with fluctuations in the use of this strategy with a low of 6.3 repeats/100 syllables in the third quarter to a high of 12.4 repeats/100 syllables in the

second quarter (net decrease of 45.0%).

6. 3. 4 Self-correction



Graph 14: Mean self-correction/100 syllables of research, control and native groups

From all the repair measures, G1 relied on self-correction to maintain their utterances. In the first quarter, G1 used this form of repair 0.89 self-corrections/100 syllables. This increase over the year to reach 3.27 self-corrections/100 syllables by the final quarter (an overall increase of 267%). Data provided on the amount of self-correction used by the control group was averaged around 2.05 self-corrections/100 while the native group used this measure predictably less at 0.57 self-corrections/100.

7. Discussion

How did the fluency measures show student progress in their fluency of English?

Utterance fluency improved dramatically over the semester. By applying this triad of

fluency measures, the results clearly support the claim that as the learners in both research groups, G1 and G2, developed confidence to express themselves with the successful integration of the TPP framework, they also progressed in fluency. These results of the research group would be in line with expectations from a previous pilot paper (Pipe & Tsushima, 2021) that as students gained more experience in the actual application of language in their paired classroom dialogues, students developed strategies to naturally process their linguistic resources in the formation, articulation, and self-monitoring stages. However, not only did students from the research outperform the control group in all categories, G1 clearly outperformed G2 and further narrowed the gap with native speakers. This must be as a result of TPP being better implemented.

Table 1: Comparison in speed performance between the former, present and native research groups over the year

		Former Research Group (G2)		Present Research Group (G1)		Native Group (NG)
	AIM	Net percentage change (%)	Net average	Net percentage change (%)	Net average	Net average
SR (syllables/min)	INCREASE	21.9	79.9	40.4	107.2	227.7
AR (syllables/min)	INCREASE	5.5	159.5	19.5	164.2	267.5
MLoR words/run (G2) syllables/run (G1)	INCREASE	17.9	3	34.2	4.6	16.9
PhonRat (% of speech)	INCREASE	16.3	49.5	19.0	65.1	85.0

Unsurprisingly, G1 performance in all speed measures (table 1), SR, AR, MLoR and PhonRat, improved. As mentioned earlier (refer to 6.1) CG performed the worst in each category with only slight gains in the speed measures as they struggled to maintain their utterances. One could possibly claim that application of the TPP framework improved speed with a more careful balance between meaningful negotiation to acquire language (Fiorella & Mayer, 2015) on the one hand, and sufficient repetitive practice on the other to reduce the level of working memory used in the activities (Ellis, 2005) orchestrated in classroom management. In fact, with more experience in applying TPP, one can see a greater increase in the net changes in the speed performance measures of G1 (e.g. SR 40.4% increase) compared to G2 net changes (e.g. SR 21.9% increase). This has been as a result of a broad range of factors. From the teacher's perspective, providing more concise

instruction, better modeled examples, more informed corrective feedback, and more successfully applied online materials could have made students more aware of speech production expectations. From the student's perspective, choosing better topics, providing more opportunities for students to apply corrective feedback, encouraging greater awareness of meta-learning strategies such as self-efficacy and perseverance, or pushing for deeper self-observation in proficiency may have led to an alteration in motivation from extrinsic to intrinsic participation which resulted in students wanting to actually speak more in class activities. However, looking at the speed measures of NG such as SR at 227.7 syllables/min (G1: 107.2 syllables/min; G2: 79.9 syllables/min) or AR at 267.5 syllables/min (G1: 164.2 syllables/min; G2: 159.5 syllables/min), there is still clear room for improvement in all categories.

Table 2: Comparison in pause performance between the former, present and native research groups over the year

		Former Research Group (G2)		Present Research Group (G1)		Native Group (NG)
	AIM	Net percentage change (%)	Net average	Net percentage change (%)	Net average	Net average
PauseRat (% of speech)	REDUCE	−13.9	50.5	−25.8	34.9	15.0
NCB Length (seconds)	REDUCE	12.0	0.88	−17.8	0.64	0.52
NCB Frequency (pause/100 syllables)	REDUCE	−8.3	25.1	−34.3	13.6	16.9
BCB Length (seconds)	REDUCE	1.7	1.09	−34.3	0.77	0.57
BCB Frequency (pause/100 syllables)	REDUCE	−6.0	21.1	−28.3	15.7	85.0

Drawing attention towards pausing (table 2), there was clear improvement for G1 but not for G2 which, at times, resembled the results of CG (refer to 6.2). With improved cognitive processing (Derwing, Munro, & Thomson, 2008; Segalowitz, 2003; Segalowitz & Freed, 2004) and faster reaction time in the testing (e.g., Ammar, 2008; Lyster & Izquierdo, 2009), it was expected that there would be less pausing. However, the net fall in PauseRat by G1 group was 25.8% which was a greater reduction compared to G2 which decreased by 15.8% and CG which dropped by 12.1%. This alteration in pausing would indicate possible improvement in the use of working memory as learners gain better access to their L2 language competence (Hao & Othman, 2020). Again, due to more finely

tuned implementation of TPP, one can see a greater reduction in pausing.

However, to indicate a shift from strenuous serial processing to more efficient parallel processing, one needs to observe a reduction in the frequency of pausing. Looking at pause length and frequency within the sentence, there was a reduction in non-clausal boundaries. Concentrating on G1's data, there was a drop of 17.8% in length and in frequency of 34.3%. By contrast, G2 only managed to reduce to a less degree the frequency of NCB by 8.3% and actually increased the length of this type of pause by 12%. This would indicate the challenges faced by G2 in attempting to become less dysfluent. Despite improvement in frequency, this group still had to take more time to process information. Spoken output by G1, on the other hand, certainly seemed less dysfluent in nature (Nakatsuhara, 2014) due to the reduction in pause duration as well as frequency in NCB. Such reductions obviously affect the prosodic boundaries between clauses in spontaneous speech (Choi, 2003; Ferreira, 1993; Horne et el., 1995).

However, for there to be a shift from serial processing to more efficient parallel processing, one must observe less frequent pausing between clausal boundaries (BCB) due to improved conceptualization (Lambert et al. 2017). Looking closer at pause location at BCB, G1 outperformed G2 with net decreases of 34.3% in the mean length of pause and 28.3% in frequency. Similar to NCB, G2 managed to reduce the frequency of BCB by 6% but increased the length slightly by 1.7% due to the strain of maintaining their utterances in the testing. In fact, this increase in frequency of BCB by G2 would suggest that this group probably reached "critical points" of processing difficulties associated with L2 speech (Segalowitz, 2010: 9) while the G1 managed to reduce this aspect of pausing as a result of better efficient parallel processing and being increasingly more automatic in accessing their declarative knowledge of syntactic, lexical and phonological rules (Kormos, 2006; Mora & Levkina, 2017). In fact, these results could also suggest that more efficient parallel processing had a knock on effect by reducing the strain on working memory resources and subsequently allowing G1 students to subsequently use more cognitive resources for allocating attention to formulating linguistic forms accurately (Fukuta, 2016, Lambert et al., 2021). As a result, G1 attained greater proficiency in their fluency than G2 due to more efficient parallel processing.

Although there was improvement in BCB and NCB, due to concerns regarding the operationalizing and reliable measuring of fluency (Housen et al. 2012), this paper also looked at repair as this would indicate improved control over L2 knowledge (Kormos, 2006; Lambert et al. 2017; Saito et al., 2018; Skehan, 2016; Skehan & Shum, 2017; Tavakoli

Table 3: Comparison in repair performance between the former, present and native research groups over the year

		Former Research Group (G2)		Present Research Group (G1)		Native Group (NG)
	AIM	Net percentage change (%)	Net average	Net percentage change (%)	Net average	Net average
Filled Pauses (/100 syllables)	REDUCE	5.9	10.2	−38.7%	8.5	2.53
False Starts (/100 syllables)	REDUCE	21.4	5.49	−72.4	0.53	267.5
Repeated Words (/100 syllables)	REDUCE	−45.0	9.3	−28.0	3.9	16.9
Self-correction (/100 syllables)	REDUCE	N/A	N/A	267.0	2.24	0.57

& Wright, 2019). This would also capture the cognitive challenges students faced in continuing their language production i.e. issues of clarity within the message produced and the strategies used by the speaker to buffer their utterances when encoding a speech plan. Unsurprisingly, the control group struggled to maintain their utterances due to a lack of development in their lexical, grammatical or phonological resources (refer to 6.3). G1, however, would seem to be more determined to apply their linguistic resources to maintain their spoken monologues. This observation would seem to reflect the data as this group relied less on filled pauses (decreased by 38.7%), false starts (decreased by 72.4%) and repeats (decreased by 28.0%), but a heavy surge in self-correction (increased by 267%). In fact, this trend in self-correction is to be expected as they learnt how to express themselves and accurately fine-tune their message while committing themselves to the conversations. Rather than pausing, these students were prepared to make mistakes in conversation and develop their lexical, grammatical, or phonological resources. G1 would seem to stretch their linguistic resources in an effort to become more proficient in attaining a more accurate message when formulating their thoughts and opinions.

Data on G2, on the other hand, would suggest less progress was made in an effort to become more cognitively fluent in expressing their utterances. There were rise in the levels of filled pauses (increase by 5.9%) and false starts (increase of 21.4%) but a reduction in repeated words (decrease of 45%). Unfortunately, no data was analysed in regards to self-correction as this was initially considered in the research plan but would have better indicated the cognitive challenges students face in continuing their language production. However, with an increase on filled pauses and false starts to maintain an

increased rate of speech production, one can conclude the students in G2 being cognitively challenged and, therefore, less fluent.

Overall, the data on repair would complement the data on speed and pausing. There was noticeable improvement in fluency in their spoken monologues with reductions in terms of speed, pause and repair due to less strain in the cognitive demands of speech production for G1 and, to a less extent, G2 (Derwing et al., 2009; Pipe & Tsushima, 2021; Tavakoli *et al*., 2020). Compared to these research groups, CG's performance would reflect lower leveled L2 speakers as they were less fluent in terms of speed and pausing, and dysfluent in their speak production in terms of frequency of pausing and repair (Kormos, 2006; Mora & Levkina, 2017; Pipe & Tsushima, 2021, Segalowitz, 2010). On the other hand, students from the research groups showed marked improvement in their English proficiency. Through effective preparation, practice, and testing in TPP, students were clearly re-orientated to conversational tasks (Pipe & Tsushima, 2021/2022). In fact, as they built strategies to maintain conversation through practice in TPP, they also started to gain confidence and became more proficient in exploring and experimenting in their spoken language discourse and thus less perturbed when challenged to express themselves – consequently becoming more apt by also producing lengthier and more complex sentences which can only indicate greater proficiency in their English abilities. Development of these sociolinguistic and metacognitive habits were, therefore, reflected in the recorded data with increased speed, reduced pausing and overall reduction in repairs. One can, therefore, support the claim that there was improved fluency as spoken language produced at an increasing speed with relative ease and less hesitation (Tavakoli et al., 2020).

8. Discussion and Recommendations

8. 1 Data elicitation method

Possibly the performance of students could partly have been affected by the procedure of the testing. As described in 5.2, students orientated themselves in cultivating their language resources and developing their speech production by generating conversation from their prepared 20 questions. In the recordings, however, students had to maintain a one-minute monologue on what they did at the weekend. These recordings may not completely reflect the complete realities of TPP in the classroom. In TPP, students pushed for much meaning negotiation by developing their socio-interlanguage

systems. The recording provided, however, were individual performances based on the same topic in which students were simply asked to explain what they did over the weekend. Although the recordings convey features of fluency, the level of commitment to the recording would have been affected by how interesting their weekend actually was. If the student had done relatively little, it would have been more challenging to maintain a monologue for a minute – especially if they were repeating the same type of monologue each week because they were limited in what they could do due to the restriction caused by COVID-19. If the recordings analysed were of a similar nature to the TPP sessions by allowing students to talk in pairs, the dynamics of the interaction would have changed and possibly have reflected greater fluency as students would be more motivated to commit to their conversation. There would have been more of a spark in the interaction and thereby lead to a more in-depth display in fluency, leading to, for example, possibly more repair to maintain conversation, increased AR and longer pausing between clauses. Instead, students may have reached a comfortable plateau and felt less inclined to stretch their lexical resource in the recorded data compared to spoken output in the class activities. As a result, despite the encouraging results, students may have in fact provides more complacent versions of their thoughts in order to reach their target to speak for a one-minute recording.

8. 2 Correction challenges

A challenge during TPP testing was still the level of corrective feedback to provide students (Pipe & Tsushima, 2021). Pragmatic accuracy and fluency follow two different constructs, one that involves acquiring pragmatic knowledge while the other one requires gaining automatic control in processing this knowledge in real time (Kasper, 2001). Although through carefully applied corrective feedback during the testing stage of TPP, students showed greater performance in their accuracy in the speaking tests, the degree and variety of corrective feedback was varied so as to not make a trade-off with performance fluency (Skehan, 1996) as this also had to be nurtured. Furthermore, it was also a concern that some students might have become too reliant on teacher involvement to ensure that there was a certain amount of progress made. For students to pursue greater fluency in their conversations, it is necessary for students to feel greater confidence in their English abilities.

The research has so far focused on a quantitative analysis of the theory of TPP. This framework needs to reflect the true understandings of the actual participants for

corrective feedback to become more effective. To enable a deeper and more genuine understanding that reflects a more true and fair description of views held by the students (Howe, 2004), a qualitative investigation could provide clarity from the students' perspective as to what could be considered a suitable level of corrective feedback. From qualitative analysis of surveys or interviews, one might begin to determine a more realistic interpretation of the cause and effect in the hypothesis between TPP and fluency by establishing a more effective approach in the application of corrective feedback through the testing stages of TPP. At the moment, it is clear that corrective feedback needs to consider concerns about accuracy level and level of reliance on teacher input. In this research, it was felt that some students would not have benefited as much as previously envisaged with corrective feedback as there was too much to consider at the formulation stage of speech production (Kormos, 2006). As a result, feedback was used more sparingly and targeted to particular aspects of grammar.

9. Conclusion

TPP shows huge potential for teachers to include in their EFL classroom. Following from the pilot paper by Pipe & Tsushima (2021), TPP framework not simply encourages students to genuinely engage in their paired conversation to improve their proficiency in English but invigorates them in their learning of English. Most of the findings of the present study have established a strong cause-effect relationship between the application of the TPP framework and improvement in fluency. Over the academic year, it is clearly evident that students can become more fluency in their speech despite the challenges Japanese students face when conversing in their English due to inappropriate pedagogic methodologies introduced at high school and cultural linguistic concerns. This is of great significance as this demonstrates that students at university can re-orientate themselves to improving their proficiency in speech production.

Appendices on Fluency Composite Measures

1. 1 Speed

Table 4: Formulae for Speech Rate, Articulation Rate, Phonation-time Ratio and Mean Length of Run

Speech Rate (SR) (syllables/min)	$\frac{\text{Total number of syllables produced from the entire narrative}}{\text{The total time (in minutes) required to produce the speech sample}}$
Articulation Rate (AR) (syllables/min)	$\frac{\text{Total number of syllables produced from the entire narrative}}{\text{The total time required to produce the speech sample excluding pause time of 300ms or above}}$
Phonational-time Ratio (PhonRat) (%)	$\frac{\text{Length of actual time spoken}}{\text{Time taken to produce the narrative}} \times 100$
Length of Runs (MLoR) (syllables/utterance)	Average mean of all utterances between pauses of 300ms or above of $\frac{\text{Number of syllables in each utterance run}}{\text{Total Utterance}}$

1. 2 Pausing

Table 5: Pause-time Ratio

Pause Ratio (PauseRat) (%)	$\frac{\text{Length of total pauses}}{\text{Time taken to produce the narrative}} \times 100$

Table 6: Formulae to determine the mean length and frequency of NCN and BCB

	Non-Clausal Boundaries (NCB)	Between-Clausal Boundaries (BCB)
Mean length (secs)	$\frac{\text{Total length of non-clausal pause}}{\text{The frequency of non-clausal pauses}}$	$\frac{\text{Total length of clausal pause}}{\text{The frequency of clausal pauses}}$
Frequency (per 100 syllables)	$\frac{\text{Total number of non-clausal pause}}{\text{100 syllable utterance}}$	$\frac{\text{Total number of clausal pause}}{\text{100 syllable utterance}}$

1. 3 Repair

Table 7: Filled Pauses, False Starts, Repeats and Self Corrections

Filled Pauses Frequency (per 100 syllables)	False Starts Frequency (per 100 syllables)	Repeats Frequency (per 100 syllables)	Self Corrections Frequency (per 100 syllables)
$\frac{\text{Total number of filled pauses}}{\text{100 syllable utterance}}$	$\frac{\text{Total number of false starts}}{\text{100 syllable utterance}}$	$\frac{\text{Total number of repeats}}{\text{100 syllable utterance}}$	$\frac{\text{Total number of self corrections}}{\text{100 syllable utterance}}$

References

Ahangari, S., Rahbar, S., & S. E. Maleki, 2015. Pronunciation or listening enhancement: Two birds

with one stone. *International Journal of Language and Applied Linguistics, 1*（2）, 13–19.

Ahmadian, M. J., & M. Tavakoli, 2011. The effects of simultaneous use of careful online planning and task repetition on accuracy, fluency, and complexity of EFL learners' oral production. *Language Teaching Research, 15*, 35–59.

Ammar, A., 2008. Prompts and recasts: Differential effects on second language morphosyntax. *Language Teaching Research*, 12, 183–210.

Banks, S., 2016. Behind Japanese students' silence in English classrooms. *Accents Asia*, 8（2）, 54–75.

Bannert R., Botinis A., Gawronska B., Katsika A., & E. Sandblom, 2003. Discourse structure and prosodic correlates. Proceedings of *the XVth International Congress of Phonetic Sciences*; Barcelona, Spain. August 3–9, 2003; Adelaide: Causal Productions; 2003, 1229–1232.

Boersma, P., & D. Weenink, 2014. *Praat: Doing phoetics by computer*［Computer program］. Version 5.4, retrieved 15 August, 2019 from http://www.praat.org/

Bortfeld, H., Leon, S. D., Bloom, J., E., Schober, M., F., & S. E. Brennan, S. E. 2001. Disfluency Rates in Conversation: Effects of Age, Relationship, Topic, Role, and Gender. Language and Speech, 44（2）, 123–147.

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

Bot, K. D., 1992. Applied linguistics. Applied Linguistics, 13（1）, 1–24.

Bowyer, D. S., 2019. Exploring the effects of recursive conversations on L2 learner beliefs. In P. Clements, A. Krause, & P. Bennett（Eds.）, *Diversity and inclusion*. Tokyo: JALT.

Brown, H. D., 2014. *Principles of language learning and teaching*. White Plains, NY: Pearson Education.

Buchanan, T. W., Laures-Gore, J., & M. Duff, 2014. Acute stress reduces speech fluency. *Biological Psychology*, 97（1）. DOI: 10.1016/j.biopsycho.2014.02.005

Butler, Y. G., 2015. English language education among young learners in East Asia: A review of current research（2004–2014）. *Language Teaching, 48*（03）, 303–342.

Butler, Y. G. & M. Iino, 2005. Current Japanese Reforms in English Language Education: The 2003 Action Plan. *Language Policy*, 4（1）, 25–45.

Choi, J. Y., 2003. Pause length and speech rate as durational cues for prosody markers. *Journal of the Acoustical Society of America*, 114（4）, 2395.

Clark, H. H. & T. Wasow, 1998. Repeating words in spontaneous speech. *Cognitive Psychology, 37*（3）, 201–242

Cutler, A., & T. Otake, 1994. Mora or phoneme? evidence for language-specific listening. *Journal of Memory and Language*, 33, 824–844. doi: 10.1006/jmla.1994.1039

de Bot, K., 1992. A bilingual production model: Levelt's 'speaking' model adapted. *Applied Linguistics*, 13, 1–24.

de Freitas, P. G., & G. A. Frangiotti, 2020. The Implicit-Explicit Dichotomy in Language Teaching/Learning: A Proposal for an Update. *International Journal of Language and Linguistics*,

7 (4), 25–38. doi: 10.30845/ijll.v7n4p4

de Groot, A. M. B., 1992. Determinants of word translation. Journal of Experimental Psychology Learning, Memory and Cognition 18, 1001–18.

de Groot, A. M. B., 1995. Determinants of bilingual lexicosemantic organization. *Computer Assisted Language Learning*, 8 (2–3), 151–180.

de Jong, N. H., 2016. Predicting pauses in L1 and L2 speech: The effects of utterance boundaries and word frequency. *International Review of Applied Linguistics in Language Teaching*, *54*, 113–132.

de Jong, N. H., 2018. Fluency in second language testing: Insights from different disciplines. *Language Assessment Quarterly*, *15*, 237–254.

de Jong, N. H., Groenhout, R., Schoonen R., & J. H. Hulstijn, 2015. Second language fluency: speaking style or proficiency? Correcting measures of second language fluency for first language behavior. *Applied Psycholinguistics*. 36 (2) : 223–243. DOI: 10.1017/

de Jong, N., & C. A. Perfetti, 2011. Fluency training in the ESL classroom: An experimental study of fluency development and proceduralization. *Language Learning, 61, 533–568*. doi: 10.1111/j.1467- 9922.2010.00620.x

de Jong, N. H., Steinel, M. P., Florijn, A., Schoonen, R., & J. H. Hulstijn, 2012. The effect of task complexity on functional adequacy, fluency and lexical diversity in speaking performances of native and non-native speakers. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 121–142). Amsterdam: John Benjamins.

Derwing, T. M., Diepenbroek, L. G., & J. A. Foote. 2012. How Well Do General-skills ESL Textbooks Address Pronunciation? TESL Canada

Derwing, T. M., Munro, M. J., & R. I. Thomson. 2008. A longitudinal study of ESL learners' fluency and comprehensibility development. *Applied Linguistics*, 29 (3), 359–380.

Derwing, T. M., Munro, M. J., Thomson, R. I., & M. J. Rossiter, 2009. The relationship between L1 fluency and L2 fluency development. Studies in Second Lan- guage Acquisition, 31, 533–557.

Doe, T., 2017. Oral fluency development activities: A one-semester study of EFL students. Thesis, Temple University Graduate Board. Retrienved from https://scholarshare.temple.edu/bitstream/handle/20.500.12613/2792/TETDEDXDoe-temple-0225E-13136.pdf?sequence=1

Dornyei, Z., & M. L. Scott, 1997. Communication strategies in a second language: definitions and taxonomies. *Language Learning*, 47 (1), 173–210.

Education First, 2020. EF English Proficiency Index. EF Education First. Retrieved from: https://www.ef.com/wwen/epi/regions/asia/japan/

Education Testing Service, 2019. *TOEFL iBT: Test and score data summary 2019*. *ETS*, United States. Reviewed from: https://www.ets.org/s/toefl/pdf/94227_unlweb.pdf

Ellis, R., 2005. *Planning and Task Performance in a Second Language*. Benjamins, Amsterdam, Netherlands, 2005.

Ellis, R., & G. Barkhuizen, 2005. *Analysing Learner Language*. Oxford: Oxford University Press.

Fiorella L., & R. E. Mayer, 2015. *Learning as a Generative Activity*. Cambridge University Press,

New York, NY, USA.

Flege, J. E., & O. S. Bohn, 1989. An instrumental study of vowel reduction and stress placement in Spanish-accented English. *Studies in Second Language Acquisition*, 11, 35–62.

Fletcher, J., 1987. Some micro and macro effects of tempo change on timing in French. *Linguistics*, 25, 951–967.

Fokes, J., Bond, Z. S., & M. Steinberg, 1984. Patterns of English word stress by native and non-native speakers. In M. P. R. van den Broecke & A. Cohen (Eds.), *Proceedings of the Tenth International Congress of Phonetic Sciences* (pp. 682–686). Dordrecht: Foris.

Foster, P., Tonkyn, A. & G. Wigglesworth, 2000. Measuring spoken language: A unit for all reasons. *Applied Linguistics*, 21, 354–75.

Fukuta, J., 2016. Effects of task repetition on learners' attention orientation in L2 oral production. *Language Teaching Research*, 20, 321–340. doi: 10.1177/1362168815570142

García-Amaya, L., 2015. A longitudinal study of filled pauses and silent pauses in second language speech. *The 7th Workshop on Disfluency in Spontaneous Speech* (DiSS 2015).

Gilquin, G., & S. De Cock, 2011. Errors and disfluencies in spoken corpora: Setting the scene. International Journal of Corpus Linguistics. 16 (2), 141–172.

Götz, S., 2013. Fluency in Native and Nonnative English Speech. Studies in Corpus Linguistics, 53. Amsterdam; Philadelphia: John Benjamins Publishing Company.

Grabe, E., & E. L. Low, 2002. Durational variability in speech and the rhythm class hypothesis. In C. Gussenhoven & N. Warner (Eds.), *Laboratory Phonology VII* (pp. 515–546). Berlin: Mouton de Gruyter.

Gráf, T., 2017. Repeats in Advanced Spoken English of Learners with Czech as L1'. *Auc Philologica*, 2017 (3), 65–78. https://doi.org/10.14712/24646830.2017.34

Hao, M., & R. Othman, 2021. Automation of Function Assignment in the Models of Speech Production and Second Language Acquisition. *Hindawi Education Research International*, 2021, 1–9. Article ID 2441598, https://doi.org/10.1155/2021/2441598.

Harmer, J., 2001. The Practice of English Language Teaching. London: Longman.

Ho Van Han, M. A., 2015. A survey of English major juniors' active versus passive learning styles at BVU. *International Journal of Information Research and Review*, 2 (3), 553–555.

Hofstede, G., Hofstede, G. J., & M. Minkov. 2010. *Cultures and Organizations: Software of the Mind* (Rev. 3rd ed.). New York: McGraw-Hill.

Horne, M., Strangert, E., & M. Heldner, 1995. Prosodic boundary strength in Swedish: final lengthening and silent interval duration. Proceedings of the *XIIIth International Congress of Phonetic Sciences*; Stockholm: KTH and Stockholm University; 1995, 170–173.

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

Housen, A., Kuiken, F., & I. Vedder, 2012. Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA (Vol. 32). Amsterdam: John Benjamins.

Howe, K. R., 2004. A critique of experimentalism. *Qualitative Inquiry*, 10 (1), 42–61.

Huensch, A., & N. Tracy–Ventura, 2017. L2 utterance fluency development before, during, and

after res- idence abroad: A multidimensional investigation. *Modern Language Journal, 101*, 275–293.

Hunter, A. M., 2017. Fluency development in the ESL classroom: The impact of immediate task repetition and procedural repetition on learners' oral fluency. (Unpublished doctoral dissertation). In P. Tavakoli, F. Nakatsuhara, & A. M. Hunter, *Aspects of Fluency Across Assessed Levels of Speaking Proficiency*. The Modern Language Journal. 104 (1), 169–191.

Japan Times, 2019. English level at Japan's secondary schools falls short of government target. *Japan Times*, April 16th, 2019. Retrieved from https://www.japantimes.co.jp/news/2019/04/16/national/english-level-japans-secondary-schools-falls-short-government-target/

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

Kasper, G., 2001. Four perspectives on L2 pragmatic development. *Applied Linguistics*, 22, 502–530.

Kawamura, K., Kudo, T., & E. M. Hail, 2006. Ways to activate students' utterances. *Ritsumeikan Gengo Bunka Kenkyu*, 18 (1), 169–181.

Kikuchi, K., & C. Browne, 2009. English education Policy for high schools in Japan: Ideals vs Reality. *RELC Journal*, 40 (2), 172–191.

Kindt, D., & D. S. Bowyer, 2018. A working paper exploring the effects of recursive conversations on participants' fluency development in a first-year EFL oral communication course. *Nagoya University of Foreign Studies Ronshu*, 2, 103–129.

Koike, Y., 2014. Explicit Pronunciation Instruction: Teaching Suprasegmentals to Japanese Learners of English. In N. Shoda & Krause (Eds.), JALT2013 Conference Proceedings. Tokyo: JALT.

Koponen, M., & H. Riggenbach, 2000. Overview: Varying perspectives on fluency. In H. Riggenbach (Ed.), Perspectives on fluency (pp. 5–24). Ann Arbor, MI: University of Michigan Press.

Kormos, J., 2006. *Speech production and second language acquisition*. Mahwah, NJ: Lawrence Erlbaum Associates.

Lambert, C., Kormos, J., & D. Minn, 2017. Task repetition and second language speech processing. *Studies in Second Language Acquisition*, 39, 167–196. Doi: 10.1017/s0272263116000085.

Lambert, C., Aubrey, S., & P. Leeming, 2021. Task Preparation and Second Language Speech Production. *TESOL Quarterly 55(2), 331–365*. DOI: 10.1002/tesq.598

Levelt, W. J. M., 1989. *Speaking from intention to articulation*. Cambridge, MA: MIT Press.

Linck, J., A., Osthus, P., Koeth, J., T., & M. F. Bunting, 2014. Working Memory and Second Language Comprehension and Production: A Meta-analysis. *Psychonomic Bulletin & Review*, 21 (4), 861–83. DOI: 10.3758/s13423-013-0565-2

Løfsgaard, K. A., 2015. The history of English education in Japan: Motivations, attitudes and methods. (Masters), University of Oslo.

Longauerová, R., 2016. *Sociophonetic study of dysfluent behaviour in native English speakers* (Un-

published BA thesis). Charles University, Prague. Available online: https://dspace.cuni.cz/handle/20.500.11956/76793 (accessed on 20 July 2020)

Lyster, R., & J. Izquierdo, 2009. Interactional feedback and instructional counterbalance. *Language Learning*, 59, 453–498.

Maclay H., & C. E. Osgood, 1959. Hesitation Phenomena in Spontaneous English Speech, WORD, 15: 1, 19–44, DOI: 10.1080/00437956.1959.11659682

Michel, M., 2011. Effects of task complexity and interaction in L2 performance. In P. Robinson, *Second Language Task Complexity: Researching the Cognition Hypothesis of Language Learning and Performance*. Amsterdam: John Benjamins Publishing Company, 141–174.

Mora, J., & M. Levkina, 2017. Task-based pronunciation teaching and research: Key issues and future directions. *Studies in Second Language Acquisition*, 39 (2), 381–399.

Moser, J., 2010. Using an AS-unit Complexity Benchmark to Measure Beginner Learner Oral Production in Communicative Tasks. *Osaka Shoin Women's University Proceedings*, (47).

Nakamura, S., 2010. Analysis of relationship between duration characteristics and subjective evaluation of English speech by Japanese learners with regard to contrast of the stressed to the unstressed. *Journal of Pan-Pacific Association of Applied Linguistics, 14* (1), 1–14.

Nakata, Y., 2011.Teachers' readiness for promoting learner autonomy: A study of Japanese EFL high school teachers. *Teaching and Teacher Education*, 27 (5), 900–910. ISSN 0742-051X

Nesic, I. D., & K. Hamidoivc, 2015. Teaching English grammar: Efficiency of inductive and deductive approaches students' perceptions. *Zbornik radova Filozofskog fakulteta u Prištini*, 45 (3), 189–205. DOI: 10.5937/zrffp45-9250

Nisbett, R. E, & T. Masuda, 2007. Culture and point of view. Intellectica: Revue de L'Assoiciation pour la Recherche Cognitive, 2–3: 46–47, 153–172.

Otake, T., Hatano, G., Cutler, A., & J. Mehler, 1993. Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language*, 32, 258–278.

Paul, S. S., 2017. Active and Passive Learning: A Comparison. *GRD Journal for Engineering* 2 (9), 27–29.

Pellicer-Sánchez, A., & N. Schmitt, 2010. Incidental vocabulary acquisition from an authentic novel: Do Things Fall Apart? *Reading in a Foreign Language*, 22 (1), 31–55.

Pica, T., Holliday, L., Lewis, N., & L. Morgenthaler, 1989. Comprehensible output as an outcome of linguistic demands on the learner. Studies in Second Language Acquisition, 11, 63–90.

Pipe J., & T. Tsushima, 2021. Improved Fluency through the Timed-Pair-Practice Framework. *The Asian Conference on Language 2021: Official Conference Proceedings*, ISSN: 2435–7030, March 2021.

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

Porter, D., 1999. Pronunciation. In Spolsky, B. *Concise Encyclopedia of Educational Linguistics*. Oxford: Pergamon Elsevier.

Prefontaine, Y., 2013. Perceptions of French Fluency in Second Language Speech Production. *Canadian Modern Language Review*, 69 (3), 324–348. DOI: 10.3138/cmlr.1748

Pulvermuller, F., 2002. *The neuroscience of language*. Cambridge: Cambridge University Press.

Saito, K., Ilkan, M., Magne, V., Tran, M. N., & S. Suzuki, 2018. Acoustic characters and learner profiles of low-, mid- and high-level second language fluency. *Applied Psycholinguistics*, 39, 593–617. DOI: 10.1017/S0142716417000571.

Schmitt, N., 2000. *Vocabulary in language teaching*. Cambridge: Cambridge University Press.

Segalowitz, N., 2003. Automaticity and second languages. In The, C. Doughty and M. Long (Eds.), *Handbook of Second Language Acquisition*, (pp. 382–388), Blackwell, Oxford, UK.

Segalowitz, N., 2010. The cognitive bases of second language fluency. New York: Routledge.

Segalowitz, N., 2016. Second language fluency and its underlying cognitive and social determinants. *International Review of Applied Linguistics*, 2016, *54* (2), 79–95. doi: 10.1515/iral-2016-9991

Segalowitz, N., & B. F. Freed, 2004. Context, contact and cognition in oral fluency acquisition: Learning Spanish in At Home and Study Abroad contexts. *Studies in Second Language Acquisition*, 26, 173–199.

Sheppard, C., & R. Ellis. 2018. The effects of awareness-raising through stimulated recall on the repeated performance of the same task and on a new task of the same type. In M. Bygate (Ed.), Learning language through task repetition (pp. 171–192). John Benjamins Publishing Company.

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

Sinyashina, E., 2020. 'Incidental + Intentional' vs 'Intentional + Incidental' Vocabulary Learning: Which is More Effective? *Complutense Journal of English Studies*, 28, 81–96. DOI: 10.5209/cjes.66685

Skehan, P., 2003. Task-based instruction. *Language Teaching*, *36*, 1–14.

Skehan, P., 2014. Limited attentional capacity, second language performance, and task-based pedagogy. In P. Skehan (Ed.), *Processing perspectives on task performance* (pp. 211–260). Amsterdam: John Benjamins. https://doi.org/10.1075/tblt.5.08ske

Skehan, P., 2015. Limited attention capacity and cognition: Two hypotheses regarding second language performance on tasks. In M. Bygate (Ed.), *Domains and directions in the development of TBLT: A decade of plenaries from the international conference* (pp. 123– 156). Amsterdam: John Benjamins.

Skehan, P., 2016. Foreign language aptitude, acquisitional sequences, and psycholinguistic processes. In G. Granena, D. O. Jackson, & Y. Yilmaz (Eds.), Cognitive individual differences in second language processing and acquisition (pp. 17–40). John Benjamins.

Skehan, P., & S. Shum, 2017. What influences performance? Personal style or the task being done? In L. Wong & K. Hyland (Eds.), *Faces of English education: Students, teachers and pedagogy* (pp. 29–43). London: Taylor & Francis.

Smith, C., 2004. Topic transitions and durational prosody in reading aloud: production and modeling. *Speech Communication*, 42, 247–270.

Steele, D., & Zhang, R. (2016). Enhancement of teacher training: Key to improvement of English Education in Japan. *Procedia – Social and Behavourial Sciences*, 18 (1), 55–88.

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

Szaszák, G., & A. Beke, 2015. Toward Exploring the Role of Disfluencies from an Acoustic Point of View: A New Aspect of (Dis) continuous Speech Prosody Modelling. In *Text, Speech, and Dialogue, 18th International Conference, TSD 2015*, Pilsen,Czech Republic, *September 14–17, 2015, Proceedings*, 369–377.

Tahira, M., 2012. Behind MEXT's new course of study guidelines. *The Language Teacher, 36* (3), 3–8. Retrieved from jalt-publications.org/tlt.

Takahashi, M., 2005. The Efficacy of GrammarInstructionin EFL Classesin Japan. *Kobe Shoin Graduate School of Letters PhD Dissertation*. 235pages.

Tavakoli, K., 2016. Fluency in monologic and dialogic task performance: Challenges in defining and measuring L2 fluency. *De Gruyter Mouton*. 54 (2), 133–150.

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

Tavakoli, P., Nakatsuhara, F. & A. M. Hunter, 2020. Aspects of Fluency Across Assessed Levels of Speaking Proficiency. *The Modern Language Journal*. Vol. 104 (1), 169–191.

Tavakoli, P., & P. Skehan, 2005. Strategic planning, task structure and performance testing. In R. El- lis (Ed.), *Planning and task performance in a second language* (pp. 239–277). Amsterdam: Benjamins.

Tavakoli, P., & C. Wright, 2019. *Second language speech fluency: from research to practice*. Cambridge University Press.

Thai, C., & F. Boers, 2016. Repeating a monologue under increasing time pressure: Effects on fluency, complexity, and accuracy. *TESOL Quarterly*, 50, 369–393. doi: 10.1002/tesq.232

Ting-Toomey, S., & L. C. Chung, 2005. *Understanding Intercultural Communication*. Oxford: Oxford University Press.

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.

Tukahara, N., 2002. The sociolinguistic situation of English in Japan. *Revista de Sociolinguistica*. Retrieved from http://www.gencat.cat/llengua/noves/noves/hm02tardor/internacional/a_nobuyuki.pdf

Valls-Ferrer, M., & J. C. Mora, 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/Philadelphia: John Benjamins Publishing Company.

Vicsi, K., & G. Szaszák, 2010. Using prosody to improve automatic speech recognition. *Speech*

Communication, 52 (5) :413–426. DOI: 10.1016/j.specom.2010.01.003

Wang, Z., 2014. On-line time pressure manipulations: L2 speaking performance under five types of planning and repetition conditions. In P. Skehan (Ed.), *Processing perspectives on task performance* (pp. 27–62). Amsterdam: John Benjamins.

Yenkimaleki, M., & Heuven, V.J. van (2016b). The effect of memory training on interpretation performance. *International Journal of English Language, Literature and Translation Studies*, 3 (3), 79–86.

Note

1) The first author was in charge of running the English course including the design and preparation of training materials and of writing the manuscript, while the second author speech data management and analyses.