

The Impact of Time-Constrained Drills on Japanese EFL Students' Speaking Fluency and Accuracy

Thomas HOLLAND

Abstract

Speaking English is a highly desirable skill for Japanese university students. However, high schools in Japan typically prioritize reading skills in preparation for university entrance exams, resulting in students having low willingness to communicate (WTC) combined with high anxiety when attempting to speak English. This research investigates the effects of time-constrained speaking drills on the speaking performance of Japanese university students, in terms of fluency (including WTC), grammatical accuracy, and partner interaction, while also exploring the extent to which students perceived the timed practices as useful and enjoyable. The study involved two groups of second-year university students at Kyushu Sangyo University: an experimental group ($n = 19$) and a control group ($n = 8$). Over 8 weeks, both groups performed pair-based speaking drills using dialogs the students prepared as homework. In the latter four weeks, the experimental group performed the drills under time constraints that decreased by 10% with each iteration, while the control group was not subjected to any time constraints. Both groups of students were assessed at mid-semester and end of semester on fluency, accuracy, and partner interaction. The experimental group also rated the timed practices in terms of usefulness and enjoyment. The experimental group showed statistically significant improvements in fluency and total scores ($p < .001$), as well as gains in accuracy that approached significance. By contrast, the control group exhibited no statistically significant improvements. Additionally, students in the experimental group rated the timed practices highly in terms of both usefulness (Mean=4.56) and enjoyment (Mean=4.44). These findings suggest that time-constrained speaking drills may provide a simple yet effective method of addressing WTC, enhancing fluency, and increasing engagement in the EFL classroom.

日本の大学生にとって、英語を話すことは非常に望ましいスキルですが、日本の高校では、大学入試に備えて読解力を優先させるのが一般的です。その結果、生徒はコミュニケーション意欲 (WTC) が低く、英語を話そうとすると不安が強くなります。本研究では、時間制限付きのスピーキングドリルが日本の大学生のスピーキング能力に及ぼす影響を、流暢さ (WTC を含む)、文法の正確さ、パートナーとのやり取りの観点から調査するとともに、生徒が時間制限付きの練習をどの程度有益で楽しいと認識しているかについても調査します。本研究では、九州産業大学の2年生2グループ、実験グループ ($n=19$) と対照グループ ($n=8$) を対象としました。8週間にわたり、両グループは宿題として準備したダイアログを使用して、ペアベースのスピーキングドリルを実施しました。後半の4週間、実験グループは、反復ごとに10%短縮される時間制限の下でドリルを実施しましたが、対照グループには時間制限はありませんでした。両グループの学生は、学期の途中と学期の終わりに流暢さ、正確さ、パートナーとのやり取りについて評価されました。実験グループは、時間制限のある練習の有用性と楽しさについても評価しました。実験グループは、流暢さと合計スコアにおいて統計的に有意な改善 ($p < .001$) を示し、正確さにおいても有意に近い向上を示しました。対照的に、コントロールグループでは統計的に有意な改善は見られませんでした。さらに、実験グループの学生は、時間制限のある練習の有用性 (平均 = 4.56) と楽しさ (平均 = 4.44) の両方について高く評価しました。これらの結果は、時間制限のあるスピーキングドリルが、WTCに対処し、流暢さを高め、EFL教室での関与を高めるためのサンプルでありながら効果的な方法となる可能性があることを示唆しています。

Background

Being functional in a language requires the ability to comprehend various forms of input as well as the capacity to generate meaningful output. Although many Japanese university students state that speaking English is one of the skills they would most like to acquire, they have only limited opportunities to practice and improve their spoken English, as reading is the main skill taught in Japanese secondary schools, in preparation for university entrance examinations (Hayafune, 2023). Unfortunately, this often leaves Japanese students with little willingness to communicate in oral English, as they have not been adequately prepared for second-language (L2) output and thus find speaking to be a highly stressful activity (Elliott & Vasquez, 2022). Overcoming the resulting silence of the language learning classroom is a well-established conundrum for instructors and students alike (Maher & King, 2022).

Fluency and WTC

Although fluency and WTC are distinct concepts, for the purposes of the present study they were evaluated together. WTC is a measure of a student's willingness to initiate speech when presented with the opportunity to do so (Kang, 2005), while fluency is broadly defined as the ability to produce smooth speech “without undue pausing” (Thompson, 2017). Prior research has found evidence of a two-way interaction between WTC and fluency of speech, such that the more fluent students are, the more likely they are to be willing to speak, and that higher levels of WTC lead to more speaking opportunities and thus improved fluency. The opposites are also true, where low WTC negatively impacts fluency and low fluency diminishes a student's WTC (Nematizadeh, 2019). The time-constrained speaking drills described in this paper attempted to address both fluency and WTC by prompting students to begin speaking without hesitation, overcoming low WTC, and encouraging sustained output for the duration of the activity, promoting fluency. The degree to which students achieved these outcomes was measured jointly under the label of “fluency.”

Individual factors affecting student WTC

Elahi Shirvan et al. (2019) conducted a meta-analysis of 22 WTC studies, involving a total of 4,794 L2 English-language learners, and found that three key variables influence foreign/second language learners' WTC: motivation, which is the desire, attitude, and effort a student applies toward acquiring the target language; language anxiety, which specifically and negatively impacts a student's ability to perform a language-related task; and perceived communicative competence, which is a student's self-belief in his or her own ability to communicate. Of these three variables, perceived communicative competence was found to exert the largest effect upon WTC.

This has implications when it comes to speech rate, as Yurtbaşı (2015) described the importance of proper volume, intonation, and particularly tempo when outlining the qualities of competent speakers. While cautioning against speaking overly fast, the researcher noted that positive correlations had repeatedly been found between the rate of speech and perceived communicative competence. When speaking faster, students' confidence in their speaking ability increased.

The impact of sociocultural factors and linguistic competence upon WTC

In addition to individual factors, Macintyre (2007) found that the factors impeding L2 WTC were multifaceted, including the sociocultural situation, and even “genetic influences

handed down from one generation to the next,” such as “intergroup climate” (pp.567–568). “Intergroup climate” refers to the concept that student performance is not down solely to individual motivation, but rather that students inherit the attitudes of the culture into which they are born. This suggests that if the students’ society holds negative attitudes towards, in this case, Japanese speakers of English, their language learning motivation and achievement are likely to be impacted (Macintyre, 2007).

Macintyre et al. (1998) also noted that learners may inherit personality traits such as extraversion, agreeableness, conscientiousness, and emotional stability, which naturally predispose them for or against language learning and L2 WTC. While acknowledging that there are variations within every population, they noted that different groups were likely to exhibit different levels of particular characteristics. For example, on the whole, American learners tend to display higher levels of extraversion in comparison to Japanese learners (Macintyre et al., 1998).

Japan in particular is frequently described as a “high-context society” where communication relies upon context and shared cultural knowledge rather than the spoken word (Albertson, 2020, p.47). Albertson (2020) notes that “face-saving and silence are important and individuality is downplayed in Japanese culture,” thus students’ reluctance to speak can reflect their desire to “preserve group harmony and avoid being the ‘odd one out’ who speaks up” (p.47). Other research has demonstrated that perceptions of peer engagement significantly impact motivation among Japanese students. Collaborating in pairs or small groups helps foster a sense of cohesion and commitment, as students are no longer studying in isolation, but working together with their peers. This can exert a significant positive effect upon motivation and L2 learning, as students feel engaged in a collaborative activity (Tanaka, 2022).

Existing research into time-constrained drills

The results of existing research into time constraints relative to performance are mixed. Thai and Boers (2015) studied 20 Vietnamese EFL students and found that fluency was markedly enhanced as students were placed under increasing time pressure. Conversely, Areta and Purwanti (2021) found that in 28 Indonesian EFL learners, “limited time makes [learners] feel hurried, anxious, panicked, and even forget the material they have prepared” (p.60).

Boers (2014) studied ten students (Five from Vietnam, one from Japan, two from Malaysia, and two from Saudi Arabia) with intermediate to advanced proficiency in English

(self-reported IELTS scores ranged from 5 to 7.5). The students chose two topics they felt comfortable talking about, such as family, hobbies, hometown, etc., and performed monologues three times. Five of the students (Boers does not identify their nationalities) spoke under 4/3/2-minute decreasing time constraints, while the remaining five spoke under fixed 3/3/3-minute time constraints. The results indicated a statistically significant increase in speech rate for the decreasing-time group relative to the fixed-time group, but also a statistically significant number of errors. Boers (2014) concluded that while excessive time pressure may negatively impact accuracy, the combination of repetition and the shrinking time condition was highly effective for fluency development. Furthermore, three of the participants preferred the shrinking time format because it was an engaging, game-like activity.

Fluency and accuracy in time-constrained speaking

Thanyawatpokin and Vollmer (2018) observed that the discussion of benefits and drawbacks regarding time-constrained speaking drills was far from settled, noting that while Nation (1989) argued that 4-3-2 would lead to improvements in accuracy, Boers (2014) found the opposite to be true. In their study, Thanyawatpokin and Vollmer (2018) studied 10 Japanese first-year high school students and first-year university students over 3 consecutive weeks in an English communication class held once per week, with student TOEFL scores ranging from 375 to 450. In the study, the students were recorded on digital voice recorders as they completed 3 sets of monologues on common topics, such as travel, under shrinking time, constant time, and expanding time conditions, rotating to a new partner for each iteration within the set. A new topic was introduced for each set and students were provided 2 minutes to prepare before beginning each set.

Thanyawatpokin and Vollmer (2018) found that the shrinking time condition led to faster speech with fewer dysfluencies, while the constant time condition also improved fluency. However, the researchers could not determine whether fluency gains resulted from time pressure or simple repetition. The expanding time condition increased overall speech production but also resulted in more dysfluencies and a slower speech rate. In terms of accuracy, the shrinking time condition showed a significant increase in mean total accuracy, however accuracy decreased significantly when examined minute-by-minute. The constant time and expanding time conditions showed a decrease in mean total accuracy, however accuracy gains were shown in the expanding time condition when examined minute-by-minute. Because of these “vastly contradictory results,” the researchers could not draw

definitive conclusions with regard to accuracy, but noted that teachers “should expect positive results when looking at fluency in the shrinking time condition” (p.259).

Speaking drills using dialogs versus monologues

Whereas existing research into time-constrained speaking exercises (such as 4/3/2) studied monologues, the research presented in this paper utilized dialogs. Previous research comparing L2 learners delivering dialogs versus monologues without time constraints found greater fluency in terms of speed, length of pause and repair measures in dialogic tasks as compared to monologic tasks (Tavakoli, 2016). However, the literature reviewed for this study found no instances of time-constrained practices utilizing dialogs, with the possible exception of research by Campbell (2013), involving a “4/3/2 speaking activity” in which a student “speaks to one other student who is required to do nothing but listen” (p.113).

Student attitudes toward timed speaking drills

Inada (2022) surveyed 93 first-year students from a Japanese university to assess student levels of enjoyment and anxiety in 14-week English courses. Results showed that “enjoyment, not anxiety, was significantly associated with improvement in the students’ English proficiency” (p.74). The research presented here theorized that the inclusion of game-like timed speaking drills might be viewed as both more useful and more enjoyable than untimed drills, and student surveys were utilized to capture students’ subjective opinions of the activities.

Aims

Following up on Thanyawatpokin and Vollmer’s study (2018), the present research challenged EFL students to speak at faster than usual rates by performing pair-based speaking drills under shrinking time conditions, to investigate the impact of time pressure upon WTC/fluency, accuracy, and partner interaction. Of particular interest was whether increases in WTC/fluency, if any, would be accompanied by decreases in accuracy.

The author’s hypothesis was that speaking drills performed under time constraints would result in statistically significant improvements in fluency over those of the non-time-constrained (control) group, with minimal loss of grammatical accuracy, and minimal reduction in partner interaction. Additionally, it was hypothesized that the decreasing time

constraints would be perceived as a game-like component and be well-received by students, supporting earlier research by Boers (2014).

Sampling and Methods

The study involved two groups of Japanese second-year university students in separate classes learning identical material from the textbook *Go on Speaking 3* (McAuliffe, 2024), taught by the study's author. Prior to the study, students' Visualizing English Language Competency (VELC) test scores ranged from 509 to 536, equivalent to 450-520 on the TOEIC test. As the VELC test does not include speaking or writing components, these scores merely provided context regarding students' general comprehension of English and were not used to evaluate students at the conclusion of the study. Course enrollments, which were beyond the control of this research, resulted in an unbalanced design, with the experimental group consisting of 19 students (11 male, eight female), and the control group consisting of eight students (four male, four female).

Students in both classes performed dialogic speaking drills once per week for four weeks, followed by a dialogic speaking test in week 5. Students then did another four weeks of dialogic speaking drills, followed by another dialogic speaking test in week 10. The textbook provided two or three question prompts for each week's speaking activity. Both classes utilized a flipped classroom model, with students studying the textbook material and preparing for the speaking drills prior to the next class by writing short (approximately 80-150 word) answers to the prompts in their textbooks as homework. Their answers followed the style of model answers provided in the book and became the dialogs used in their speaking drills. The dialog topics varied by week, and covered subjects common in daily conversation (Appendix A), such as health and lifestyle, travel, and spending money.

As part of their homework, students were directed to memorize and practice speaking the dialogs they had written, either alone or with a friend, classmate, or family member. Then in class, prior to the first drill, students were also provided five minutes to silently review their dialogs, make any changes they saw fit, and memorize them as thoroughly as possible. At the end of weeks 4 and 9, students' homework also included reviewing the grading sheets and rating guidelines in their textbooks, provided in Japanese and English, in preparation for the speaking tests in weeks 5 and 10, respectively (Appendix B, McAuliffe, 2024).

To perform the drill, students stood in lines, facing each other, to form pairs. If there was an odd number of students, the instructor joined the activity, taking the role of a student.

One student in each pair (Student A) asked the question prompts from the textbook, while the other student (Student B) delivered the dialog he or she had prepared. At the conclusion of each dialog, students rotated partners, until the drill was performed four times. In both control and experimental groups, students were permitted to refer to the prepared dialogs they had written in their textbooks, if necessary, during the first two deliveries. For the final two drills, they exchanged textbooks and were encouraged to speak from memory, although their partners could provide hints if necessary.

Time-constrained speaking drills

The control group was not time-constrained during the speaking drills, while the experimental group completed the drills within decreasing amounts of time during the second half of the semester (weeks 6-9). For the experimental group, the amount of time decreased by 10 percent with each drill iteration, as shown on a large digital clock projected on a screen in the front of the classroom.

In order to encourage fluency, the instructor directed the experimental group to concentrate on their messages rather than grammatical forms (Nation, 2013). The instructor also emphasized that attempting to complete the pair dialogs within the time limits was simply a fun challenge, rather than a strict requirement, and that although they would have to stop when time ran out, no penalty would be assessed for not finishing within the time limit.

Base time for time-constrained drills

Because the four weekly topics for the time-constrained speaking drills (Appendix A, topics 5-8) required answers of different lengths, using the same, fixed times each week, such as 4/3/2- minutes, was not tenable. To address this, the number of words in the textbook's two model answers for topics 1 through 4 were counted, then both control and experimental classes were timed during weeks 1-4 to determine the mean amounts of time it took student pairs to finish conversing about each respective topic. The mean number of words-per-second the student pairs had spoken was calculated to be 0.88, or 53 words per minute, as shown in Table 1. This included a small amount of time spent switching from Student A being the primary speaker to Student B.

The number of words in the mean of the model answers for topics 5 through 8 were then divided by 0.88 (mean words per second or WPS), to arrive at the "base time" for the time-constrained speaking drills, as shown in Table 2. ("Base time" being the longest interval, or what would be the "4" in "4/3/2".) Each subsequent drill was then set to be 10,

20, and 30-percent faster than the base time. At the fastest, if their prepared dialogs contained the same number of words as the mean model answers, students would need to speak at a rate of 68 words per minute (30% faster than the base time of 53 WPM) to finish within the time limit. For reference, native speakers speak “at around 150 words per minute” (Nation, 2013, p.36). Actual student speaking rates were not measured.

Table 1

Mean Words Per Second Spoken by Student Pairs Without Time Constraints

| | Topic 1 | Topic 2 | Topic 3 | Topic 4 | Mean WPS (Model Ans A & B, Topics 1-4) |
|----------------------------------------------------------------------------|---------|---------|---------|---------|----------------------------------------------|
| Mean seconds for student pairs to complete speaking (not time-constrained) | 133 | 151 | 185 | 150 | |
| # of words in Model Ans A | 111 | 110 | 163 | 178 | |
| # of words in Model Ans B | 103 | 96 | 161 | 166 | |
| Mean # of words in both Model Ans | 107 | 103 | 162 | 172 | |
| Mean words per second (WPS) | 0.80 | 0.68 | 0.88 | 1.15 | 0.88 |

Table 2

Base Time for Time-Constrained Drills

| | Topic 5 | Topic 6 | Topic 7 | Topic 8 |
|----------------------------------------------------------------------------------------|---------|---------|---------|---------|
| # of words in Model Ans A | 213 | 154 | 163 | 175 |
| # of words in Model Ans B | 192 | 132 | 173 | 127 |
| Mean # of words in both Model Ans | 203 | 143 | 168 | 151 |
| Base Time in secs (Mean # of words in both Model Ans divided by Mean WPS from Table 1) | 240 | 163 | 192 | 172 |

Data collection methodology

At mid-semester and the end of the semester (e.g., week 5 and week 10), the instructor evaluated the students on fluency, accuracy, and interaction, using paper-based grading sheets with 4-point Likert Scales, as specified in the *Go on Speaking 3* textbook (Appendix B, McAuliffe, 2024). Two students at a time were directed to join the instructor at the back of the classroom, with all three persons sitting at desks arranged in a triangular formation. The instructor then asked one student (Student A) the questions from one of the dialogs that had been practiced (Appendix A), with the instructor taking the role of Student

B. At the conclusion of the dialog, the instructor then asked the other student a different question from the dialogs that had been practiced, and again the instructor took the role of Student B.

No time constraints were placed upon students' responses. This was done to determine if students in the experimental group had made actual improvements over the previous four weeks of timed practices, so that their speaking ability would now exhibit measurable gains without the presence of a timer. Each student was rated on fluency and accuracy, with fluency scores being negatively impacted for students who exhibited lengthy pauses or challenges in WTC, such as lengthy delays in starting to speak. The two students were then instructed to have a discussion together on a general conversation topic, with the instructor taking the role of an observer. For example, "Please talk about music" or "Please talk about Tokyo." Again, no time constraints were placed upon students' responses, and the instructor stopped the conversation after the students were judged to have discussed the topic in sufficient detail or appeared to have run out of things to say. Student interaction was rated based upon each student's ability to ask and respond appropriately and effectively in communication with his or her partner.

In week 10, after the speaking tests but before receiving their scores from the instructor, students in the experimental group also rated the degree to which they felt the timed practices had been useful and enjoyable, relative to the untimed practices. Ratings were captured using paper-based surveys with 6-point Likert scales, and questions were written in both Japanese and English (Appendix C).

The data collected from the instructor-conducted evaluations on fluency, accuracy, and interaction were analyzed using a paired samples t-test to compare the control group to itself, in weeks 5 and 10. Similarly, a paired samples t-test was used to compare the experimental group to itself, in weeks 5 and 10. A Mann-Whitney U test was then used to compare the instructor-conducted evaluations from the control and experimental groups against each other in weeks 5 and 10. This non-parametric version of the independent samples t-test test was utilized because of the large difference between the number of students in the two groups, as is appropriate for situations where sample sizes differ greatly (Urdan, 2010). Finally, student ratings of usefulness and enjoyment of the timed practices over the untimed practices were reported and interpreted using descriptive statistics.

Analysis

Comparison within groups

Within the control group, the negative t values in the paired samples t-test, shown in Table 3, indicated some improvement in scores from Test 1 to Test 2, however these were not statistically significant (defined as $p < .05$), and the effect sizes (Cohen's d) ranged from no effect to medium effect size, according to L2 learning benchmarks outlined by Plonsky and Oswald (2014), where “ d values in the neighborhood of .40 should be considered small, .70 medium, and 1.00 large” (Plonsky & Oswald, 2014, p.889). Mean scores for all variables indicated only minor or no improvements from Test 1 to Test 2, as shown in Table 4.

Table 3

Control Group. Comparison of Test 1 Results to Test 2 Results

| Test 1 Results | Test 2 Results | t | df | p | Cohen's d | SE Cohen's d |
|--------------------|--------------------|--------|------|-------|-------------|----------------|
| Test 1 Fluency | Test 2 Fluency | -1.000 | 7 | 0.351 | -0.354 | 0.361 |
| Test 1 Accuracy | Test 2 Accuracy | -2.049 | 7 | 0.080 | -0.725 | 0.298 |
| Test 1 Interaction | Test 2 Interaction | 0.000 | 7 | 1.000 | -0.000 | 0.325 |
| Test 1 Total | Test 2 Total | -1.210 | 7 | 0.265 | -0.428 | 0.331 |

Note. Student's t-test.

Table 4

Control Group. Comparison of Mean Scores on Test 1 and Test 2

| | N | Mean | SD | SE | Coefficient of variation |
|--------------------|---|-------|-------|-------|--------------------------|
| Test 1 Fluency | 8 | 2.375 | 1.188 | 0.420 | 0.500 |
| Test 2 Fluency | 8 | 2.750 | 0.886 | 0.313 | 0.322 |
| Test 1 Accuracy | 8 | 2.625 | 0.744 | 0.263 | 0.283 |
| Test 2 Accuracy | 8 | 3.000 | 0.535 | 0.189 | 0.178 |
| Test 1 Interaction | 8 | 3.000 | 0.535 | 0.189 | 0.178 |
| Test 2 Interaction | 8 | 3.000 | 0.926 | 0.327 | 0.309 |
| Test 1 Total | 8 | 8.000 | 1.927 | 0.681 | 0.241 |
| Test 2 Total | 8 | 8.750 | 1.982 | 0.701 | 0.227 |

For the experimental group, the negative t values in the paired samples t-test, shown in Table 5, indicated statistically significant gains in fluency, accuracy, and total ($p < .001$ for all three variables), along with medium to large effect sizes (Cohen's $d = .952, .920, 1.216$, respectively) for those variables (Plonsky & Oswald, 2014).

There was no significant change in partner interaction. Mean scores for all variables indicated improvements from Test 1 to Test 2, as shown in Table 6.

Table 5

Experimental Group. Comparison of Test 1 Results to Test 2 Results

| Test 1 Results | Test 2 Results | <i>t</i> | <i>df</i> | <i>p</i> | Cohen's <i>d</i> | SE Cohen's <i>d</i> |
|--------------------|--------------------|----------|-----------|----------|------------------|---------------------|
| Test 1 Fluency | Test 2 Fluency | -4.150 | 18 | < .001 | -0.952 | 0.272 |
| Test 1 Accuracy | Test 2 Accuracy | -4.009 | 18 | < .001 | -0.920 | 0.289 |
| Test 1 Interaction | Test 2 Interaction | -2.388 | 18 | 0.028 | -0.548 | 0.308 |
| Test 1 Total | Test 2 Total | -5.299 | 18 | < .001 | -1.216 | 0.280 |

Note. Student's *t*-test.

Table 6

Experimental Group. Comparison of Mean Scores on Test 1 and Test 2

| | N | Mean | SD | SE | Coefficient of variation |
|--------------------|----|--------|-------|-------|--------------------------|
| Test 1 Fluency | 19 | 2.974 | 0.754 | 0.173 | 0.254 |
| Test 2 Fluency | 19 | 3.605 | 0.542 | 0.124 | 0.150 |
| Test 1 Accuracy | 19 | 2.921 | 0.534 | 0.122 | 0.183 |
| Test 2 Accuracy | 19 | 3.500 | 0.645 | 0.148 | 0.184 |
| Test 1 Interaction | 19 | 2.947 | 0.705 | 0.162 | 0.239 |
| Test 2 Interaction | 19 | 3.368 | 0.496 | 0.114 | 0.147 |
| Test 1 Total | 19 | 8.842 | 1.573 | 0.361 | 0.178 |
| Test 2 Total | 19 | 10.474 | 1.264 | 0.290 | 0.121 |

Comparison between groups

In Table 7, analysis revealed that on Test 2 fluency, there was a statistically significant difference in the scores of the groups ($p = .016$), with mean fluency rising from 2.974 to 3.605 in the experimental group as compared to a more modest increase of 2.375 to 2.750 in the control (see Table 8, below). Effect size for Test 2 fluency was also large ($r_b = .553$), with effect sizes of $r_b < 0.3$ considered small, $r_b 0.3$ to 0.5 medium, and $r_b > 0.5$ large (DATAtab Team, 2025). This difference was not present on Test 1 Fluency ($p = .286$), suggesting that the time-constrained speaking drills had made an impact.

There was also a statistically significant difference in the Test 2 total score ($p = .034$) with a large effect size ($r_b = .520$) (DATAtab Team, 2025). Mean accuracy scores also improved, as shown in Table 8 below, increasing from 2.921 to 3.500 in the experimental group versus a lesser improvement in the control group of 2.625 to 3.000, although this failed to reach statistical significance ($p = .066$) with medium effect size ($r_b = .421$) (DATAtab Team, 2025). Test 2 partner interaction showed no significant improvement ($p = .316$).

Table 7

Control Group Versus Experimental Group. Comparison of Test 1 and Test 2 Results

| | U | p | Rank-Biserial Correlation (r_b) | SE Rank-Biserial Correlation |
|--------------------|--------|-------|-------------------------------------|------------------------------|
| Test 1 Fluency | 56.500 | 0.286 | -0.257 | 0.243 |
| Test 2 Fluency | 34.000 | 0.016 | -0.553 | 0.243 |
| Test 1 Accuracy | 55.500 | 0.227 | -0.270 | 0.243 |
| Test 2 Accuracy | 44.000 | 0.066 | -0.421 | 0.243 |
| Test 1 Interaction | 79.500 | 0.856 | 0.046 | 0.243 |
| Test 2 Interaction | 58.500 | 0.316 | -0.230 | 0.243 |
| Test 1 Total | 55.500 | 0.281 | -0.270 | 0.243 |
| Test 2 Total | 36.500 | 0.034 | -0.520 | 0.243 |

Note. Mann-Whitney U test. Effect size is given by the rank biserial correlation (r_b).

Table 8

Mean Scores of Control Group Versus Experimental Group on Test 1 and Test 2

| | Group | N | Mean | SD | SE | Coefficient of variation | Mean Rank | Sum Rank |
|--------------------|---------|----|--------|-------|-------|--------------------------|-----------|----------|
| Test 1 Fluency | control | 8 | 2.375 | 1.188 | 0.42 | 0.500 | 11.563 | 92.500 |
| | exp | 19 | 2.974 | 0.754 | 0.173 | 0.254 | 15.026 | 285.500 |
| Test 2 Fluency | control | 8 | 2.750 | 0.886 | 0.313 | 0.322 | 8.750 | 70.000 |
| | exp | 19 | 3.605 | 0.542 | 0.124 | 0.150 | 16.211 | 308.000 |
| Test 1 Accuracy | control | 8 | 2.625 | 0.744 | 0.263 | 0.283 | 11.438 | 91.500 |
| | exp | 19 | 2.921 | 0.534 | 0.122 | 0.183 | 15.079 | 286.500 |
| Test 2 Accuracy | control | 8 | 3.000 | 0.535 | 0.189 | 0.178 | 10.000 | 80.000 |
| | exp | 19 | 3.500 | 0.645 | 0.148 | 0.184 | 15.684 | 298.000 |
| Test 1 Interaction | control | 8 | 3.000 | 0.535 | 0.189 | 0.178 | 14.438 | 115.500 |
| | exp | 19 | 2.947 | 0.705 | 0.162 | 0.239 | 13.816 | 262.500 |
| Test 2 Interaction | control | 8 | 3.000 | 0.926 | 0.327 | 0.309 | 11.813 | 94.500 |
| | exp | 19 | 3.368 | 0.496 | 0.114 | 0.147 | 14.921 | 283.500 |
| Test 1 Total | control | 8 | 8.000 | 1.927 | 0.681 | 0.241 | 11.438 | 91.500 |
| | exp | 19 | 8.842 | 1.573 | 0.361 | 0.178 | 15.079 | 286.500 |
| Test 2 Total | control | 8 | 8.750 | 1.982 | 0.701 | 0.227 | 9.063 | 72.500 |
| | exp | 19 | 10.474 | 1.264 | 0.290 | 0.121 | 16.079 | 305.500 |

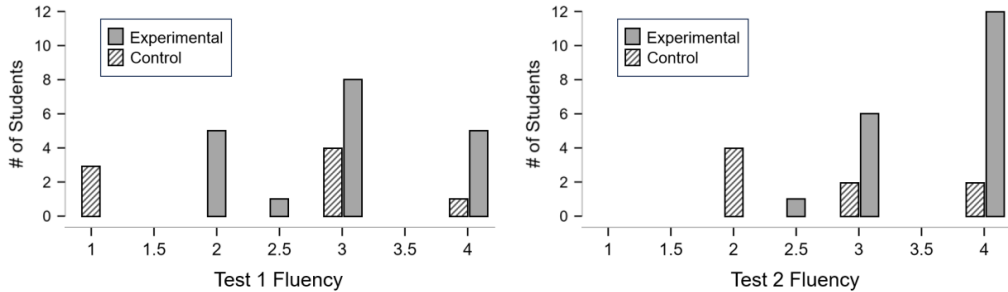
Distribution of fluency and total scores

Figures 1 and 2 compare the distribution of student fluency and total scores in the groups on Tests 1 and 2. Figure 1 illustrates significant improvements in fluency in the experimental group, where 12 students received the highest rating of four on Test 2, a considerable increase over Test 1 where five students received the same rating. By contrast, fluency in the control group improved primarily on the lower end of the scale, where three

students scored a one on Test 1. By Test 2, no students in the control were rated at one, and four students were rated at two. The number of students who received a three had decreased from four to two, indicating that while one student had improved from three to four, another student had done more poorly, going from three to two.

Figure 1

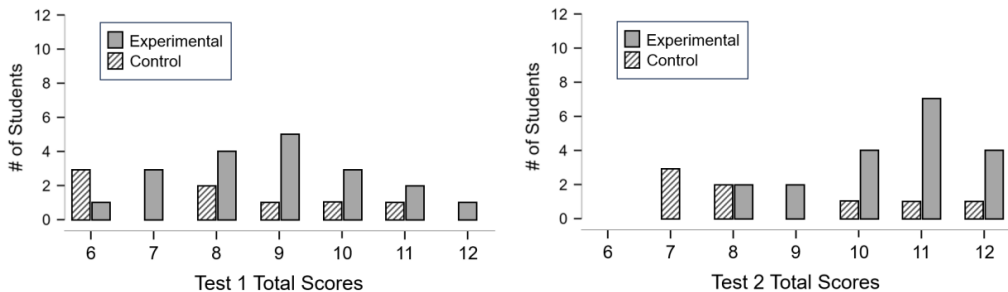
Fluency of Control Group Versus Experimental Group on Test 1 and Test 2



In Figure 2, Test 1 shows a bell-shaped distribution of total scores within the experimental group, while the control group is weighted toward the lower end of the scale, where three students received the lowest possible total rating of six. On Test 2, the experimental group showed clear improvement, with 15 of the 19 students scoring between 10 and 12. The control group made only slight improvements, with scores still overweighted on the low end of the scale.

Figure 2

Total scores of Control Group Versus Experimental Group on Test 1 and Test 2



Usefulness and enjoyability of timed practices

Figure 3 indicates that the majority of students (10 students) in the experimental group rated the timed practices as five out of six in terms usefulness over untimed practices

(with six being the most useful), and another six students rating them four out of six. The mean rating for all students was 4.556, as indicated in Table 9. Note that one student did not submit the survey, thus the number of respondents (N) was 18 rather than 19.

In terms of enjoyment, Figure 4 indicates that the majority of students (eight students) in the experimental group rated the timed practices as five out of six versus the untimed practices (with six being the most enjoyable), and another seven students rating them four out of six. Here, the mean rating for the 18 students responding was 4.444, as shown in Table 9.

Figure 3

Experimental Group, Week 10. Degree to Which Students Agreed with the Statement That Timed Practices Were More Useful Than Untimed Practices

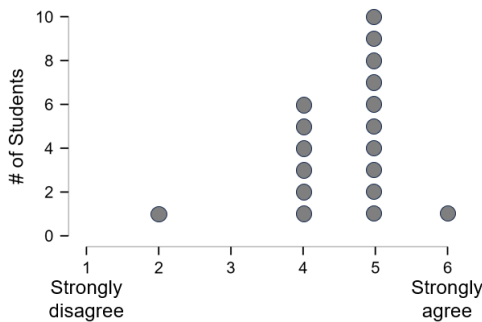


Figure 4

Experimental Group, Week 10. Degree to Which Students Agreed with the Statement That Timed Practices Were More Enjoyable Than Untimed Practices

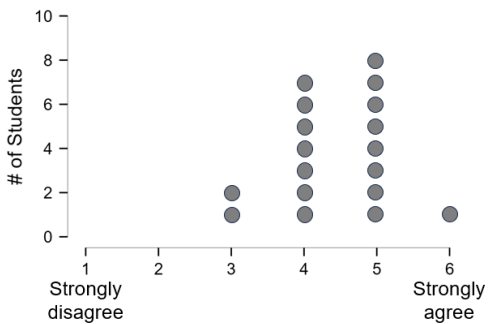


Table 9

Experimental Group, Week 10. Degree to Which Students Agreed with Statements That Timed Practices Were More Useful and More Enjoyable Than Untimed Practices

Descriptives

| | N | Mean | SD | SE | Coefficient of variation |
|----------------------|----|-------|-------|-------|--------------------------|
| Timed more useful | 18 | 4.556 | 0.856 | 0.202 | 0.188 |
| Timed more enjoyable | 18 | 4.444 | 0.784 | 0.185 | 0.176 |

Discussion

The statistically significant improvements in fluency, accuracy, and total score within the experimental group (Table 5) suggest that timed speaking practices were effective for improving students' speaking performance. The lack of significant changes in all measured variables within the control group (Table 3) further reinforces the conclusion that the observed changes were due to the timed practices rather than simply repetition alone. Furthermore, comparing the groups to one another (Table 7) revealed a statistically significant difference in fluency and total score, indicating that the timed practices were effective in prompting students to begin speaking immediately and continuing to speak at a heightened rate, with minimal production challenges related to WTC, and that the gains in fluency and total score were likely attributable to the time-constrained speaking drills.

Utilizing timed practices to overcome barriers in WTC

To understand how the timed speaking drills in the present study may have exerted a positive influence upon students' fluency and WTC (Table 5), it is worth considering the timed drills in light of the three variables of language anxiety, motivation, and perceived communicative competence (Elahi Shirvan et al., 2019), as well as the sociocultural environment of the typical Japanese classroom.

In terms of language anxiety, the game-like element of trying to beat the clock encouraged spontaneous output with a minimum of forethought. Focusing on the goal of completing the drills within the time limits rather than on grammar forms may have played a role in directing student attention away from self-consciousness, thereby reducing speech anxiety and positively impacting WTC. Because the same drill was repeated four times, students were also relieved of the burden of having to speak perfectly on the first and only try. Additionally, all student pairs spoke concurrently, so that no student would feel the pressure of "being the 'odd one out' who speaks up" (Albertson, 2020). Quite the contrary,

failing to begin speaking promptly would have singled a student out, and disadvantaged their speaking partner, thus the use of a timer might be considered one method of addressing WTC challenges in the classroom setting.

The timed pair-work practices created a unified, collaborative environment by involving all students rapidly interacting with four of their classmates as they rotated through partners. This promoted peer engagement and motivated students to join in as part of the group, reducing WTC impediments (Tanaka, 2022). Having all members speaking at once made this both a pair exercise as well as a group activity involving the class as a whole, encouraging participation by tapping into the Japanese cultural norm of preserving social harmony (Albertson, 2020).

Finally, in terms of perceived communicative competence, encouraging students to speak more rapidly may have boosted their self-perceptions of themselves as competent speakers, which in turn increased their WTC (Yurtbaşı, 2015). Overcoming students' reluctance to speak is no small feat, and getting them to speak rapidly while boosting their self-confidence as speakers of English is nothing short of remarkable.

In terms of practical implications, time constraints can be applied to many classroom speaking activities where the goal is to overcome low WTC and promote fluency (and possibly accuracy, as discussed below). Repeatedly performing the same drill under decreasing time limits adds a game-like element to what might otherwise be a boring and repetitive activity. It is of course important to retain the game-like quality of the exercise, and not to use the timer to simply push students to work harder. Maintaining an encouraging and harmonious classroom environment is essential to promoting output.

The results of this study serve to reinforce how effective gamified activities are for learning, and that creating them does not have to be complicated. Timed practices can certainly be applied to language skills other than speaking, with different outcome goals. It is also worth considering what other simple game-like elements could be added to routine classroom activities, to take them from dull to productive and engaging.

The effect of time constraints on accuracy

The current study sheds light on the question of whether improvements in accuracy were the result of time constraints or simply repetition. The results showed not only a much clearer improvement in accuracy within the experimental group ($p < .001$), but also a notable improvement in accuracy in the experimental group over the control group (from $p = .227$ on Test 1 to $p = .066$ on Test 2). Although this fell short of statistical significance, the effect size

was medium to large (Cohen's $d = .811$) and suggests that the improvement was the result of the intervention (Plonsky & Oswald, 2014).

Student attitudes toward timed practices

As shown in Figures 3 and 4 above, results of the survey suggest that students viewed the timed practices quite positively, both in terms of usefulness and enjoyability, relative to the untimed practices. This was further supported by the instructor's observation that the students appeared livelier and more engaged while attempting to meet the time challenges than they had been in weeks 1-4. This provides support for the study's hypothesis that game-like elements, namely time constraints, would improve students' perception of learning activities (Boers, 2014), and highlights the difference that something seemingly minor, such as a clock, can make upon student attitudes towards learning.

Differences from previous studies

The current research differs from previous studies in three significant ways. First, the time constraints allowed students relatively more time to speak when compared with other studies. In the present study, to finish within all time limits, the fastest speed students needed to speak at was a mean rate of 68 words per minute, as noted in the description of Table 2. This was a comparably leisurely pace, relative to 8 case studies reviewed by Nation (1989), where speaking rates ranged from 84 to 196 words per minute. It is possible that the accuracy improvements noted in the present study were attributable to students being provided with relatively more time in which to speak, such that students were pushed enough to improve their fluency and WTC, but not to the point where accuracy would suffer.

Secondly, while previous research was conducted on monologues, the current study utilized dialogs. This was done based upon the author's belief that dialogs more closely represent situations students are likely to encounter in real life, such as face-to-face conversations, as opposed to monologues, which are limited in their utility to situations like lectures, public speaking, and English class.

Thirdly, unlike previous studies, the data in this research were not collected during the speaking practice drills, but rather on subsequent tests a week later, without time constraints. Arguably, this indicates that the data from the experimental group on Test 2 were the result of tangible, internalized improvements, and that the four weeks of timed practice had actually made students more fluent speakers.

Limitations

This study has several limitations that should be considered, the first being that no baselines were established for fluency, accuracy, or interaction at the start of the semester, in week 1. This would have enabled the comparison of variables at three points in time, rather than the existing two, in weeks 5 and 10.

This study would have also benefitted from the implementation of a crossover design, so that the current control group performed timed speaking practices in weeks 1 through 4, followed by untimed practices in weeks 6 through 9 (the experimental group would stay the same).

The fact that students were rated by only one evaluator, the study's author, was also a notable limitation. Although every effort was made to be fair and impartial, bias, or the appearance thereof, may diminish the results of the study. Future studies would be well served by utilizing a rater independent from the study, or ideally, multiple independent raters.

Actual student speaking rates were not captured, which presents an opportunity for future research. It would be valuable to explore the relationship between actual speech rates and time constraints of varying durations, to determine if more optimal timespans could be established, ones which could result in greater positive impacts upon WTC/fluency and partner interaction with minimal reductions in accuracy.

Additionally, the sample size was relatively small, particularly in the control group of only 8 students, which may limit the generalizability of the results. There were also several external factors that were not controlled for, including classes for the two groups being held at different times of day, individual personalities within the groups, and even the weather on practice and test days. Future research could improve upon this by utilizing similar methodologies with larger sample sizes, in classes held at different times throughout the week, to diminish the impact of confounding variables.

It is also worth noting that the fluency and total scores of two groups were already divergent from Test 1, prior to the intervention (as evident in Figures 1 and 2), with 3 learners in the control group scoring the lowest on each respective scale. It is possible that the control group's lack of significant progress on Test 2 was simply a reflection of underperformance. Because the group was not subject to the intervention, as in a crossover study, it cannot be determined if time-constrained drills would have produced the same results as in the experimental group.

Finally, the survey of student attitudes was quite brief, and the phrasing of questions displayed a positive-language bias. A longer, more robust survey containing both positive and

negative questions, with the same concepts queried through multiple question phrasings, would improve validity. It would also be valuable to capture students' subjective opinions through short essay questions. The relationship between timed practice and student engagement presents another opportunity for exploration in future research.

Conclusion

Speaking practices performed under decreasing time conditions, utilizing dialogs prepared by students, resulted in significant improvements in fluency, including WTC. Grammatical accuracy also improved, although not to a statistically significant extent, suggesting that time pressure may help focus students' attention. These results were not merely temporary, as students were assessed a week after practice sessions had finished. This study found that repetition alone, without time constraints, did not result in similar improvements.

Incorporating time constraints into speaking practices effectively served to diminish the burden of the "W" in WTC. Instead of waiting to slowly wade into conversation, the timer provided an impetus for students to begin speaking immediately. The game-like element of the on-screen clock directed student attention outward, away from self-consciousness, and toward the goal of completing within the time limit, possibly helping to reduce speaking anxiety. At the same time, students' increased rate of output likely bolstered their self-perceived communicative competence. The timer also placed students in a situation where, if they did not promptly begin and continue speaking, they would single themselves out, which is something to be avoided within the sociocultural environment of Japan. Some or all of these factors likely contributed to increased WTC and improved fluency within the time-constrained group.

While encouraging students to speak faster certainly has limits (too fast and output becomes unintelligible), judiciously applied timed speaking drills offer one avenue for improving students' self-perceived competence and subsequently WTC/fluency. Moreover, time constraints add an element of fun. Even setting aside effects upon fluency or accuracy, time-constrained drills may be beneficial in classrooms where students appear apathetic or are contending with WTC issues.

Prompting students to stand and speak with one another is a foundational activity for many language courses, and the addition of time constraints with decreasing time intervals

provides a simple way to gamify classroom speaking practices by focusing students on the task and challenging them to beat the clock.

References

- Albertson, B. P. (2020). Promoting Japanese university students' participation in English classroom discussions: Towards a culturally-informed bottom-up approach. *Journal of Pan-Pacific Association of Applied Linguistics*, 24(1), 45–66. <https://doi.org/10.25256/PAAL.24.1.3>
- Areta, D. P., & Purwanti, I. T. (2021). Factors contributing to EFL learners' speaking performance. *International Journal of Educational Best Practices*, 5(1), 60–79. <https://doi.org/10.31258/ijebp.v5n1.p60-78>
- Boers, F. (2014). A reappraisal of the 4/3/2 activity. *RELC Journal*, 45(3), 221–235. <https://doi.org/10.1177/0033688214546964>
- Campbell, M. (2013). Improving learners' oral complexity with the 4/3/2 speaking activity. *New Directions in Teaching and Learning English Discussion*, 1(2), pp.113–118. <https://doi.org/10.14992/00016168>
- DATAtab Team. (2025, February 2). *Mann-Whitney U-test*. datatab. <https://datatab.net/tutorial/mann-whitney-u-test>
- Elahi Shirvan, M., Khajavy, G., MacIntyre, P., & Taherian, T. (2019). A meta-analysis of L2 willingness to communicate and its three high-evidence correlates. *J Psycholinguist Res*, 48, 1241–1267. <https://doi.org/10.1007/s10936-019-09656-9>
- Elliott, L., & Vasquez, M. (2022). Speaking anxiety: Japanese students in the EFL environment. *The Asian Conference on Education 2021: Official Conference Proceedings*, 453–464. ISSN: 2186-5892. <https://doi.org/10.22492/issn.2186-5892.2022.35>
- Hayafune, Y. (2023). How vocabulary knowledge and reading skills affect speaking performance in story retelling test. *Language Education & Technology* 60 (2023), 97–120. Retrieved from https://www.jstage.jst.go.jp/article/let/60/0/60_97/_pdf/-char/en
- Inada, T. (2022). Levels of enjoyment in class are closely related to improved English proficiency. *English Language Teaching*, 15(5), 69–76. <https://doi.org/10.5539/elt.v15n5p69>
- Kang, S. (2005). Dynamic emergence of situational willingness to communicate in a second language. *System*, 33, 277–292. <https://doi.org/10.1016/j.system.2004.10.004>
- Macintyre, P. D. (2007). Willingness to communicate in the second language: Understanding the decision to speak as a volitional process. *The Modern Language Journal*, 91(4), 564–576. <https://doi.org/10.1111/j.1540-4781.2007.00623.x>
- Macintyre, P. D., Clement, R., Dornyei, Z., & Noels, K. A. (1998). Conceptualizing willingness to communicate in a L2: A situational model of L2 confidence and affiliation. *The Modern Language Journal*, 82, iv, 545–562. <https://doi.org/10.1111/j.1540-4781.1998.tb05543.x>
- Maher, K., & King, J. (2022). 'The silence kills me.': 'Silence' as a trigger of speaking-related anxiety in the English-medium classroom. *English Teaching & Learning*, 46, 213–234. <https://doi.org/10.1007/s42321-022-00119-4>
- McAuliffe, M. (2024). *Go on speaking 3* (1st ed.). Xreading Publishing.
- Nation, P. (1989). Improving speaking fluency. *ScienceDirect*. [https://doi.org/10.1016/0346-251X\(89\)90010-9](https://doi.org/10.1016/0346-251X(89)90010-9). Retrieved from <https://www.wgtn.ac.nz/lals/resources/paul->

- nations-resources/paul-nations-publications/publications/documents/1989-Fluency.pdf
- Nation, P. (2013). *What should every EFL teacher know?* Compass Publishing.
- Nematizadeh, S. (2019). Willingness to communicate and second language speech fluency: A complex dynamic systems perspective. *Carleton University*.
<https://doi.org/10.22215/etd/2019-13680>
- Plonsky, L., & Oswald, F. L. (2014). How big is “big”? Interpreting effect sizes in L2 research. *Language Learning*, 64(4), 878–912. <https://doi.org/10.1111/lang.12079>
- Tanaka, M. (2022). Individual perceptions of group work environment, motivation, and achievement. *International Review of Applied Linguistics in Language Teaching*, 60(4), 1201–1225. <https://doi.org/10.1515/iral-2020-0183>
- Tavakoli, P. (2016). Fluency in monologic and dialogic task performance: Challenges in defining and measuring L2 fluency. *International Review of Applied Linguistics in Language Teaching*, 54(2), 133–150. <https://doi.org/10.1515/iral-2016-9994>
- Thai, C., & Boers, F. (2015). Repeating a monologue under increasing time pressure: Effects on fluency, complexity, and accuracy. *TESOL Quarterly*, 50(2), 369–393. <https://doi.org/10.1002/tesq.232>
- Thanyawatpokin, B., & Vollmer, C. (2018). Flipping the 4-3-2 activity. *Language teaching in a global Age: Shaping the classroom, shaping the world*. Tokyo: JALT, 255-261. Retrieved from https://www.researchgate.net/profile/Carl-Vollmer-2/publication/330466779_Flipping_the_4-3-2_activity/links/5c415288458515a4c72e45fb/Flipping-the-4-3-2-activity.pdf
- Thompson, C. (2017). Defining and measuring L2 fluency. *Seinan Gakuin University Center for Language Education*. Retrieved from <http://repository.seinangu.ac.jp/bitstream/handle/123456789/1687/1e-n7v1-p1-8-tho.pdf?sequence=1&isAllowed=y>
- Urduan, T. C. (2010). *Statistics in plain English* (3rd ed.). Routledge.
- Yurtbaşı, M. (2015). Why should speech rate (tempo) be integrated into pronunciation teaching curriculum? *Journal of Education and Future*, 8, 85–101. Retrieved from <https://files.eric.ed.gov/fulltext/ED566202.pdf?form=MG0AV3>

Appendix A

Student Weekly Dialog Topics from Go on Speaking 3 (McAuliffe, 2024)

1. University Life
2. Health & Lifestyle
3. Travel
4. Clothes & Accessories
5. Spending Money [week 6]
6. Friends [week 7]
7. Skills [week 8]
8. Careers [week 9]

Appendix B

Instructor's Grading Sheet Combined with Rating Guidelines for Students, from Go on Speaking 3 (McAuliffe, 2024)

| Criteria | Score |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------|
| Part 1 | |
| Fluency: you will get a higher score if you can speak without long pauses. In addition, you will get a higher score if you can give detailed answers. | 1 - 2 - 3 - 4 |
| Vocabulary & Grammar: you will get a higher score for using appropriate grammar and vocabulary. | 1 - 2 - 3 - 4 |
| Part 2 | |
| Interaction: you will get a higher score if you participate fully in the conversation, for example, by giving detailed answers to your partner's questions, by asking appropriate questions, and by responding effectively to your partner's answers. | 1 - 2 - 3 - 4 |
| Total: for each of the above criteria, you can get a maximum score of 4 and a minimum score of 1. | |

Appendix C

Experimental Group, Week 10. Student Survey Regarding Speaking Practices

1. Overall, I think the timed speaking practices help me more [are more useful] than the untimed practices
全体として、時間制限なしのスピーキング練習よりも時間を制限したスピーキング練習の方が役立つと思います
2. I think that timed practices are more enjoyable than untimed practices
時間制限なしの練習より時間制限のある練習の方が楽しいと思います