EFL Speaking in the Japanese University Context: A Comparison of Task-Based and Presentation-Practice-Production Pedagogical Approaches

Pino Cutrone
Nagasaki University, Japan
(Email: k-pino@nagasaki-u.ac.jp)

Received: 12 March, 2025; Accepted: 20 April, 2025; Published: 26 May, 2025
https://doi.org/10.58304/tc.20250501

Abstract
66 Japanese university students (enrolled in an English Listening and Speaking course) participated in this study. The students in this study were divided into groups and given different types of instruction. Group 1 received TBLT, while Group 2 received PPP instruction. TBLT relied more on implicit instruction, which involved the instructor first helping students to prepare for a task, then facilitating their work on it, and finally guiding them in reflecting on their language use and learning (Ellis, 2021a). In comparison, the PPP approach provided explicit instruction, which involved the instructor introducing new language concepts, guiding students in controlled practice activities, and then having them use the language freely in a final production task (Case, 2008). Assessments of the English oral competence of members of both groups were conducted at the beginning and the end of their respective courses (after 15 weeks, which consisted of approximately 45 hours of instruction). Each of these assessments involved participating in a dyadic conversation with a classmate, which was transcribed verbatim and analyzed. The results demonstrate that both TBLT and PPP improved Japanese EFL students' speaking skills; however, TBLT was more effective overall. Both groups made similar gains in Accuracy; however, students who received TBLT showed marked improvement in (temporal and hesitation) Fluency, while those who received PPP instruction did not. Further, the TBLT group exhibited greater gains in Syntactic and Lexical Complexity. Consequently, the writer urges language instructors in Japan to actively explore TBLT methods in English communication classes.

Keywords
Japanese EFL university students, task-based language teaching (TBLT), presentation–practice–production (PPP), speaking skills, oral competence

Introduction

Despite expending significant resources to promote English language acquisition among its populace, Japan has struggled to produce competent speakers of English (Mizuho, 2017). Various scholars, such as Ellis (1991), Farooq (2005), and Taylor (2020), highlight the limitations of university graduates' spoken English. This phenomenon is further reflected by the term *false beginners* used in university EFL materials (Helgesen et al., 2007; Martin, 2003). These students, while possessing basic reading skills and grammar knowledge from high school, lack practical communication abilities.

There are several potential reasons for this, one of which is a failure of the instructional methods employed in Japanese EFL classrooms (Taylor, 2020). Despite ongoing calls for increased communication activities in EFL classrooms in Japan, meaningful change remains

challenging as long as the country remains entrenched in the current entrance examination washback cycle (Caine, 2005; Reesor, 2003; Sakui, 2004; Taylor, 2020). Because teachers must prepare students for non-communicative entrance exams—often emphasizing rote memorization, grammar translation, and passive knowledge of a foreign language—the instructional methods used in these classes also tend to be non-communicative, offering limited opportunities for language use. Thus, the purpose of this study is to inform language pedagogy by exploring (and comparing) two instructional approaches, Task-based Language Teaching (TBLT) and Presentation–Practice–Production (PPP) instruction, designed to improve the speaking skills of learners in the Japanese EFL university context.

Literature Review

What is Task-Based Language Teaching (TBLT)?

In brief, Task-Based Language Teaching (TBLT) is an approach to foreign language education grounded in communicative and constructivist theories of language learning. TBLT prioritizes meaningful interaction and the development of fluency through authentic tasks, which are the central focus of lessons. Lessons are designed and sequenced around these tasks, with communication viewed as a process (focusing on how language is learned) rather than a product (focusing solely on what is learned) (Nunan, 1988). While various definitions of pedagogical tasks exist, some common characteristics put forward by Skehan (1998) include the following:

1. Meaning is primary.
2. Learners are not given other people's meaning to regurgitate.
3. There is some sort of relationship to comparable real-world activities.
4. Task completion has some priority.
5. The assessment of the task is in terms of outcome. (p. 147)

Hence, while an instructor asking their students to speak about a topic (such as food) for a specified amount of time may be a task in general terms, it would not be the kind of pedagogic task Skehan (1998) referred to above. This is because simply talking about food does not necessarily require meaning to be exchanged or negotiated, nor is there a task outcome that needs to be achieved (i.e., no reason to communicate). One way to transform this into a pedagogical task is to have students select menu items they believe foreign visitors might struggle to understand and then determine how they would explain these dishes to them. Students now have a reason to communicate with one another—namely, to solve a problem, creating opportunities for the negotiation of meaning. Additionally, their communication is now goal-oriented, as they aim to describe foods that may be unfamiliar to foreign visitors.

How are TBLT Lessons and Syllabi Organized?

Having described what a *task* is above, the next step is to consider how tasks are employed in an actual lesson and across the syllabus of a course. The first thing to decipher is that *Task-Based Language Teaching* (TBLT) should not be confused with *Task-Supported Language Teaching* (TSLT). As Long, Lu, and Yi (1998) explain:

> Task-supported LT simply means use of miscellaneous pedagogic tasks unrelated to learners' real world needs to practice items in a synthetic linguistic syllabus of some kind, usually a traditional grammatical, lexical or notional-functional syllabus. Task-supported LT suffers, therefore, from most of the well-known problems (irrelevance to student needs, psycholinguistic implausibility, boring lessons, etc.) characteristic of skill-building approaches that employ synthetic linguistic syllabi and the Present-

> Practice-Produce (PPP) methodology typically employed to deliver such syllabi at the classroom level. (pp. 88-89)

Henceforward, with this in mind, it should be noted that the writer's conceptualization of TBLT is strongly task-driven, positioning tasks as the foundational element of each lesson, from which all other instructional components, such as language focus, skill development, and assessment, are derived. This contrasts with TSLT, where tasks serve a more supplementary role within a traditionally structured curriculum.

Following a traditional form of TBLT, the writer adopts Willis's (1998) eight-step framework for a task-based lesson, as follows:

Table 1
The Eight Steps of a Task-Based Lesson Proposed by Willis (1998)

<table>
<tr><td colspan="3">Step 1: Pre-Task
The teacher introduces the topic and the task to the class, emphasizes key words and phrases, and supports learners in understanding task instructions and getting ready.</td></tr>
<tr><td colspan="3">TASK CYCLE</td></tr>
<tr><td>Step 2: Task
Learners work on the task in pairs or small groups while the teacher observes from a distance, encouraging all communication attempts without correcting them. This informal setting allows students to feel comfortable experimenting, knowing that it is ok to make mistakes (fluency).</td><td>Step 3: Planning
Learners plan a report to the class, either orally or in writing, on how they carried out the task and what they decided or discovered. Since this reporting stage is public, students will naturally aim for accuracy, with the teacher available to provide language support if needed.</td><td>Step 4: Report
Some groups present their reports to the class or exchange written reports to compare results. The teacher facilitates the process and provides feedback on the content of the reports. Unlike the Task in Step 2, the Report is formal and, thus, enables learners to recycle language in different contexts.</td></tr>
<tr><td colspan="3">Step 5: Model Comparison
Learners might listen to a recording of others completing a similar task to compare approaches. Alternatively, they could read a text that is either similar to what they wrote or related to the task's topic.</td></tr>
<tr><td colspan="3">LANGUAGE FOCUS</td></tr>
<tr><td colspan="2">Step 6: Analysis
Learners analyze and discuss specific features of the text or transcript from the recording. They can add new words, phrases, and patterns to their vocabulary books, increasing their awareness of language they need to complete similar tasks in the future.</td><td>Step 7: Practice
The teacher has learners practice new words, phrases, and patterns from the data, either during or after the analysis. This helps students develop the language they will need to successfully complete similar tasks in the future.</td></tr>
<tr><td colspan="3">Step 8: Post-Task
After completing this sequence, learners may benefit from attempting a similar task with a different partner. With the new skills and strategies gained in the previous steps, they are better prepared to succeed in completing the task.</td></tr>
</table>

In the framework above, the core task of the lesson is shown as the first task in the task cycle (and directly after the Pre-task). Some language practitioners may find it confusing why the core task would be done before any models of language have been presented and/or any target language has been taught. This, however, is the essence of a task-based lesson in the sense that learners initially have the opportunity to try to achieve the goals of the task on their own. In doing so, the learner will develop a heightened awareness of the language they ultimately need to accomplish the task. Thus, subsequently, when language models are presented, language analysis exercises are completed, and students ultimately have further opportunities to accomplish tasks similar to the ones they have already tried, learners are in a much better position for success.

Moreover, another hallmark of a task-based lesson involves the task cycle. Following the core task, learners have the opportunity to prepare and subsequently give a short report (in a communication class, this would entail a short speech about the discussions and conclusions the groups arrived at in the core task). In doing so, learners can reinforce the task's target language in both informal (in the core task) and formal settings (in the speech in front of the class). Another advantage of this approach is that educators are allowing learners to focus on a multitude of skills while recycling the same language across tasks. In the *core task*, which is conducted informally in small groups, learners can focus on fluency, whereas in the *report*, learners focus primarily on accuracy because the circumstances of the communication are more formal (i.e., public, rehearsed, and final).

Lastly, lessons in a task-based syllabus should be sequenced and graded based on the complexity of the core tasks. The first lesson should feature the simplest core task, while the final lesson should include the most challenging one. Task complexity is determined not by linguistic elements (e.g., grammar and vocabulary) but by inherent task characteristics. Research has identified several cognitive dimensions that affect the level of task difficulty:

- Modality: Speaking is more demanding than writing, and listening is more demanding than reading (Ellis, 1987).
- Stakes: Tasks with significant consequences for accuracy are more difficult than those without (Willis, 1993).
- Control: Tasks where participants have more control are easier than those with less (Pica, Kanagy, & Falodun, 1993).
- Planning time: Tasks that include planning time are easier than ones without planning time (Bygate, 1987).
- Number of elements: Tasks involving more elements are more difficult than tasks involving fewer elements (Brown, Anderson, Shilcock, & Yule, 1984).
- Single versus dual tasks: An example of a single task would be the gap-fill activity commonly found in EFL speaking classes that involves one student giving directions using a map, and the other student follows them. The focus is on clear, spoken communication. This same activity could be made more difficult as a dual task by adding obstacles (e.g., roadblocks). The speaker must adjust directions in real time, and the listener must identify mistakes or ask for clarification, increasing the cognitive load (Robinson, 1998).
- Prior knowledge: Tasks in a domain which learners have prior knowledge are easier than tasks in a domain in which learners have no prior knowledge of (Robinson, 1998).

Considering these cognitive dimensions, syllabus designers may try to sequence their task-based syllabi according to task complexity; however, sometimes, decisions will have to be made accounting for thematic elements of their syllabus and course. For instance, if the theme

of a book (or a course) is studying abroad, it makes sense that tasks involving flying to and returning from a foreign country be at the beginning and the end of the book, respectively (regardless of task complexity). Despite decades of research, task complexity remains a challenge in TBLT due to difficulties in its precise assessment (Mudinillah, Rahmi, & Taro, 2024).

What is Presentation–Practice–Production (PPP) Instruction?

The Presentation, Practice, and Production (PPP) instructional method has its roots in behaviorist and structuralist approaches to language learning, which emphasize habit formation and accuracy. In particular, British linguist and educator Charles L. Fries, through his systematic presentation of language structures, made foundational contributions to structural linguistics and language teaching methods in the 1940s and 1950s, laying the groundwork for the development of PPP (Fries, 1945; Fries, 1952). Later, in his influential book *Teaching Oral English*, Byrne (1976) further developed and refined the PPP model, emphasizing the importance of the gradual progression from presenting new language items, through controlled practice, to more open-ended production tasks. His work was instrumental in making PPP a widely used methodology, particularly in English Language Teaching (ELT).

PPP, thereafter, became widespread in the 1980s as part of the Communicative Language Teaching (CLT) movement, although some critics argue that it tends to follow a more traditional, teacher-centered approach rather than a truly communicative model. Still widely used more than four decades after its emergence (Harris, 2015), it has demonstrated remarkable longevity. But what does PPP actually entail? According to Andersen (2016), the three stages of PPP are envisaged as follows:

> Presentation: language features (including grammar, lexis, and functional exponents) are selected and sequenced in advance for explicit instruction, typically involving contextualized presentation followed by elicited clarification of meaning, form, and use.
> Practice: controlled practice of the feature is provided, typically including written exercises (such as gap-fills), controlled speaking practice activities (for example 'Find someone who …'), and oral drills.
> Production: opportunities for use of the feature are provided through free production activities that attempt to simulate real-world language usage (spoken or written) such as role-plays, discussions, email exchanges, and story writing, when correction and integrated form focus can be provided by the teacher. (p. 226)

In Japan, the PPP approach, particularly its more traditional and teacher-centered form, continues to be the prevailing framework in language instruction (Sato, 2010).

How are PPP Lessons and Syllabi Organized?

In defining Presentation, Practice, Production (PPP) above, we have been able to outline, in general terms, how PPP lessons are organized. However, as Case (2016) points out, PPP has evolved in modern times and may contain many variants. That is, very few if any modern PPP materials consist of an unbroken sequence of different language points presented and practiced in succession. Instead, modern PPP—often referred to as "PPP plus skills" (Case, 2016, para 2)—frequently integrates language skills development, particularly reading and listening, alongside traditional grammar and vocabulary instruction. This method contrasts with strict PPP, where lessons are heavily structured around introducing, practicing, and producing specific grammar points in a set sequence.

Furthermore, the writer now explores the possible ways in which PPP syllabi can be sequenced. In PPP-based syllabi, lessons are carefully structured to guide learners through a logical progression of language development, with grammar often playing an essential role in this process. Grammar in a PPP syllabus is sequenced from simple to complex based on cognitive complexity, frequency of use, linguistic transparency, and communicative necessity. High-frequency, seemingly straightforward structures like the present simple are introduced first, while more complex forms, such as conditionals and passive voice, are taught later to build upon previously acquired knowledge and support both accuracy and fluency in communication.

Thematic progression ensures that new language concepts, including grammar, are introduced within relevant topics, starting with simpler structures and gradually increasing in complexity. This approach helps learners build a strong foundation while engaging with meaningful content. Cumulative learning reinforces previously taught material while incorporating new elements, allowing students to see how different grammatical structures connect and evolve. Regular review and recycling of grammar help solidify understanding and improve retention.

Additionally, clear objectives shape the sequencing of lessons, with each stage—Presentation, Practice, and Production—designed to support measurable progress. The Presentation stage introduces grammar in context, followed by controlled practice activities that focus on accuracy. Finally, the Production stage encourages learners to use these structures in real-world communication. This structured approach ensures that students not only develop grammatical accuracy but also gain the confidence to apply their knowledge effectively.

TBLT and PPP in the Research Literature

Historically, much of the TBLT research has been grounded in Second Language Acquisition (SLA) theories, such as Long's (1985) Interaction Hypothesis and Swain's (1993) Output Hypothesis, often emphasizing the cognitive and affective aspects of learning rather than real-world applications (Van den Branden, 2006). Many early studies explored theoretical frameworks, learner engagement, and task complexity rather than how teachers can actually implement TBLT in different educational settings. While TBLT is strongly supported by theory and research, its adoption in EFL contexts like Japan remains limited, overshadowed by traditional methods such as PPP (Ellis, 2006). Decades after its inception, TBLT continues to be plagued by misconceptions and a lack of clear understanding (Cutrone, 2018; Ellis, 2009).

Unlike with TBLT, the research focus of PPP has often been more practical and pedagogical than theoretical (Skehan, 1998). Early research on PPP focused on its effectiveness in structuring lessons, particularly for grammar and controlled language use. PPP research has typically been more concerned with lesson planning, classroom implementation, and short-term learning outcomes (Richards & Rogers, 2014). However, in recent years, there has been more criticism of PPP from an SLA perspective, arguing that it does not align well with theories of language acquisition that emphasize interaction and meaningful communication (Ellis, 2018). Studies have questioned whether the *practice* and *production* stages truly lead to spontaneous, long-term language acquisition or if they simply encourage rote learning (Andersen, 2017; Ellis, 2018). Some researchers have also explored hybrid approaches, integrating PPP with more communicative or task-based methods to make it more effective (Ellis, 2018, 2021b; Willis & Willis, 2007).

In their book *Task-Based Language Teaching: Theory and Practice*, Ellis, Skehan, Li, Shintani, and Lambert (2020) provide a meta-analysis of some recent studies comparing TBLT and PPP. Ellis et al. (2020) were careful not to assert that one size fits all, recognizing that many factors—

such as the specific implementation of each approach, learners' proficiency levels, learning context, evaluation methods, and experimental design—must be taken into account. With this in mind, Ellis et al. (2020) were able to shed some light on how TBLT and PPP stack up against each other in various contexts.

First, Beretta and Davies (1985), who conducted a study in India measuring the general English proficiency of 390 beginner-level secondary school students, found that their TBLT group had, overall, demonstrated superior acquisition over their PPP group. In contrast, one of the only studies definitively supporting PPP over TBLT was by Sheen (2005), who investigated how grade six French-speaking elementary students in Canada learned two grammatical structures: WH interrogatives and adverb placement. Sheen (2005) conducted three tests but only reported results from the oral interview and grammaticality judgment test. The PPP group showed improvement in both tests, while the TBLT group did not, leading Sheen to conclude that PPP was more effective than TBLT for learning the target structures.

However, as Ellis et al. (2020) have pointed out, due to some fundamental flaws in the design and writing up of the study, it is difficult to preclude any conclusions. A major issue with Sheen's (2005, p. 288) study is that while detailed descriptions of the PPP instruction were provided, the TBLT instruction was poorly outlined, offering only vague details about "enjoyable tasks and game activities." Furthermore, the TBLT lessons lacked a clear focus on form and corrective feedback, and the testing primarily favored the PPP group, failing to assess learners' ability to use the target structures in spontaneous speech.

Next, in investigating the vocabulary acquisition capabilities of 30 university students in an elementary Spanish class, De la Fuente (2006) found that the two TBLT groups (i.e., one with explicit instruction and the other without) in his study had both outpaced the PPP group. Similarly, De Ridder, Vangehuchten, and Gomez (2007) conducted a study at Antwerp University with a group of 68 intermediate-level students of Spanish (i.e., majoring in business and economics), who were assigned randomly to two groups (TBLT and PPP). Participants' general and oral performances were assessed at the end of the program based on pronunciation, fluency, intonation, sociolinguistic competence, lexical competence, and grammatical competence.

The TBLT group outperformed the comparison group in grammar, vocabulary, and social appropriateness, but, surprisingly, showed no advantage in fluency. Meanwhile, the PPP group scored higher in pronunciation and intonation. One issue with this program comparison, however, as the authors acknowledged, is that the final oral assessment differed between groups. The PPP group gave individual presentations to an examiner (a native or near-native speaker), while the TBLT group was assessed based on conversations with their partners (i.e., classmates). Furthermore, Arslanyilmaz (2013) examined the comparative effects of computer-delivered TBLT and PPP by analyzing the learners' oral CAF (Complexity, Accuracy, and Fluency) during the computer-mediated lessons. Although the study's findings did not produce statistically significant results (due to a limited sample size of 28), the TBLT group notably outperformed the PPP group in all areas.

In addition, Gonz âez-Lloret and Nielson (2015) conducted two studies in a Spanish for Specific Purposes program designed for students preparing to become border patrol agents in the United States. In the first study, which analyzed the oral CAF of 39 students (in two groups of 20 and 19 respectively), Gonz âez-Lloret and Nielson (2015) found that the TBLT group was far superior in developing oral fluency and complexity than the PPP group and just as

effective in developing grammatical accuracy. The second study, which was not a comparative one, examined the overall proficiency, sentence mastery, vocabulary, fluency, and pronunciation of 256 students of a TBLT course. The findings showed that learners had improved significantly in all categories and that the benefits of TBLT were evident at all levels of proficiency.

While the studies described above are useful in showing the benefits of TBLT in various contexts, it is useful to now discuss comparative studies that have been conducted specifically in the Japanese EFL context. First, we examine the study conducted by Shintani (2013), who investigated the vocabulary acquisition capabilities of 45 Japanese children, all aged six, with no prior experience of any L2 learning. Unlike the previously mentioned studies that contained a pre-test post-test design, Shintani included a control group to compare with her two experimental groups (i.e., a TBLT group and a PPP group). While both the TBLT and PPP groups outperformed the control group in the acquisition of nouns, the TBLT group outperformed both the PPP and the control groups in the acquisition of adjectives. As part of the same study, Shintani (2015) also examined the incidental acquisition of two grammatical features—plural *-s* and the copula *be*—neither of which was explicitly taught in the TBLT or PPP lessons. The results showed that the TBLT group acquired plural *-s* but not the copula *be*, while the PPP group did not acquire either structure.

Finally, Harris and Leeming (2021) tracked 75 Japanese university students (who were non-English majors) over a year to compare the impact of TBLT and PPP on speaking skills and self-efficacy (SE). The results indicated that students in both the TBLT and PPP classes improved in proficiency and self-efficacy, with no significant differences observed in their final outcomes. Interestingly, however, patterns of growth were different. The TBLT showed initial gains and then tapered off, whereas the PPP group demonstrated more gradual progress, with notable gains emerging in the second semester. As Leeming and Harris (2021) pointed out, the proficiency task required students to complete a narrative without any pre-supplied language, which may have advantaged the TBLT group, as it closely reflected the task-based methods practiced in their classes.

This study is particularly noteworthy as it is the first to raise the possibility that the duration of instruction may influence the effectiveness of different pedagogical approaches. It highlights an important variable that future research should explore further. Similarly, this study underscores the need for further research into how different pedagogical approaches affect learners' affective factors. While several studies have demonstrated the positive impact of TBLT on Japanese learners' willingness to communicate (WTC) and confidence (Cutrone & Beh, 2017, 2018, 2024), Leeming and Harris (2021), working in a comparable context, found that most of the students they interviewed preferred the PPP approach.

In conclusion, as discussed in this literature review, research in this area remains relatively limited. However, among the existing comparative studies, the majority, though not all, tend to favor TBLT over the PPP approach. Nonetheless, several studies have yielded incomplete and/or inconclusive results. Therefore, further research is needed not only to clarify this topic but also to address the gaps and inconsistencies in the current findings.

Research Question

The aims of this study are threefold: to inform EFL language pedagogy in Japan, to improve oral competence in English classes in Japanese universities, and to ultimately assess the efficacy of TBLT vis-a-vis PPP. The writer hypothesizes that TBLT will outperform PPP

where Fluency is concerned; however, PPP may show greater gains in terms of Syntactic Complexity and Accuracy. To this end, the following research question (RQ) has been formulated:

How do TBLT and PPP compare in affecting Japanese EFL university students' speaking skills in terms of Fluency (both temporal and hesitation), Syntactic Complexity, Accuracy, and Lexical Complexity in this study?

To answer this research question, the researcher employed a quantitative analysis, as described and reported in the following sections.

Methodology

Participants

This action research study involved 66 first-year students enrolled in two Listening and Speaking courses (from April 2023 to August 2023 and April 2024 to August 2024, respectively) at the School of Global Humanities and Social Sciences, Nagasaki University. All participants resided in Nagasaki Prefecture, Japan, and were between 18 and 20 years old. The study employed an opportunistic sampling method, with the researcher recruiting readily accessible students from two EFL courses he was teaching. Informed consent procedures were followed, and participation was completely voluntary; students were free to omit any sections or withdraw from the study at any time. To protect anonymity, participants are identified by pseudonyms throughout this research, and their privacy will always be strictly maintained. The participants had an average of eight years of English language experience, including six years of mandatory English instruction in junior and senior high school. Based on interviews with the researcher, the estimated oral proficiency of students in this study ranged from A2 to B2 on the Common European Framework of Reference for Languages (CEFR) scale.

Procedures and Data Collection Methods

This action research study involved the steps described in Table 2.

Table 2
The Four Procedural Steps Involved in this Study

Step 1	Pre-Tests (consisting of video recording dyadic conversations) were conducted within one week of Step 2 beginning.
Step 2	Treatment: Group 1 received 15 weeks (i.e., approximately 45 hours) of TBLT, while Group 2 received 15 weeks (i.e., approximately 45 hours) of PPP instruction.
Step 3	Post-Tests (consisting of video recording dyadic conversations) were conducted within one week of Step 2 ending.
Step 4	Data analysis: each of the dyadic conversations were transcribed verbatim and analyzed in terms of Fluency (temporal and hesitation), Syntactic Complexity, Accuracy, and Lexical Complexity.

The four procedural steps involved in this study are described in greater detail.

Step 1 (Pre-Test): All 66 participants in this study received identical Pre-Tests occurring within one week of each other. This consisted of observation sessions (i.e., participants engaged in nine-minute dyadic video-recorded conversations, of which the middle three minutes were used as data in this study). These conversations took place in the primary researcher's office at Nagasaki University. The video recording equipment used was a Sony digital video camera, which was placed unobtrusively in the corner of the room. While the conversation was being recorded, only the participants were present in the room. Initial conversational prompts (i.e., involving peer mentoring) were offered to help stimulate conversation; however, it was made clear to all participants that they were free to talk about anything they liked.

Step 2 (Treatment): The 66 participants of this study comprised two groups of 33 students enrolled in two separate EFL Listening and Speaking courses. From April 2023 to August 2023, one group of 33 students (i.e., to be called the TBLT group hereafter) received TBLT. Subsequently, from April 2024 to August 2024, the other group of 33 students (i.e., to be called the PPP group hereafter) received PPP instruction. Treatment for both groups began within one week of the Pre-Test and consisted of 90-minute lessons twice a week over a 15-week semester (approximately 45 hours in total).

Following the tenets of TBLT described above, the TBLT group was taught using the coursebook *Welcome to Kyushu, Japan* (Cutrone & Beh, 2015). According to Ellis (2018), this is one of the few coursebooks previously available in the Japanese EFL market that truly encompasses TBLT. In the same way, following the principles of PPP described above, the PPP group was taught using the coursebook *American Headway 2*, 3rd edition (Soars & Soars, 2016). According to Case (2008, para 3), textbook "series like *Headway* (the most popular internationally available series of textbooks ever) are often taken as the defining example of PPP that have forced the whole industry to follow them." While this coursebook is organized and sequenced according to grammatical structures, it often combines language skills development, especially reading and listening, with traditional grammar and vocabulary instruction.

Step 3: The Post-Test, which was conducted within one week of the course ending, consisted of having each student undergo the same process described in Step 1 above.

Step 4: The final step of this study was to analyze and interpret data produced by the dyadic conversations. This involved first transcribing the middle three minutes of each conversation. Subsequently, as shown in Table 3, the speech data were examined in terms of Fluency (temporal and hesitation), Syntactic Complexity, Accuracy, and Lexical Complexity. Data produced in this study were analyzed quantitatively using JASP (2024). The tables in the Results section will report the pre- and post-treatment means and standard deviations (SDs) of the two groups' performances across the categories outlined in Table 3.

Since this data met the assumptions for parametric tests, paired sample t-tests were used to determine whether the differences between the means of the two groups were significant. The alpha level was set at .05. To provide an overview, the tables in the Results section will include the mean difference (the difference between the mean of the Pre-Test and the mean of the Post-Test), mean percent difference (the average absolute percentage difference between paired values), t (a statistic indicating how many standard errors the mean difference is away from zero), p (representing the tail probability of the observed t-statistic under the assumption of the null hypothesis), and d (Cohen's d was used to calculate the effect size to detail the magnitude

of change) scores. Cohen (1988) classified effect sizes as small (d = 0.2), medium (d = 0.5), and large (d ≥ 0.8).

Table 3
Analytic Framework Adapted from Inoue (2010)

Aspect	Measures	Definition
Fluency (Temporal)	Mean length of runs	Average no. of syllables produced in utterances between pauses of 0.25 seconds and above
	Speech rate	Total no. of syllables produced in a given speech sample divided by the amount of total time required to produce the speech sample (including pause time) expressed in seconds
Fluency (Hesitation)	No. of repetitions	No. of immediate and verbatim repetition of a word or a phrase
	No. of false starts	No. of utterances that are abandoned before completion
	No. of reformulations	No. of phrases or clauses that are repeated with some modification either to syntax, morphology, or word order
	No. or replacements	No. of lexical items that are substituted for another
	No. of unfilled long pauses	No. of unfilled pauses that occur for more than .5 seconds
	No. of filled pauses	No. of pauses that include filler words (like "um," "uh," or "hmm") during the silence.
Syntactic Complexity	No. of words per AS-unit	Average no. of words per AS-unit
	No. of subordinate clauses per AS-unit	Average no. of subordinate clauses per AS-unit
Accuracy	Percentage of error-free clauses	% of clauses which do not contain any error to the total number of clauses
	No. of errors per AS-unit	No. of errors divided by the total number of AS-units
	Errors per 100 words	No. of errors divided by the total number of words produced divided by 100
Lexical Complexity	Lexical Frequency Profile Vocab Size (word list checker)	% of words listed in the LFP Vocabulary Lists (JACET 8000)
	Type-Token Ratio (TTR)	No. of types (different words) divided by the number of tokens (total number of words) in a given text

To assess Syntactic Complexity and Accuracy, it is necessary to identify Analysis of Speech Units (AS-Units). In its simplest form, an AS-Unit is roughly equivalent to a clause in a sentence. However, as Foster, Tonkyn, and Wigglesworth (2000) more specifically define, "an AS-Unit is a single speaker's utterance consisting of an independent clause, or sub-clausal unit, together with any subordinate clause associated with either" (p. 365). This framework accounts for the fragmentary and elliptical nature of spoken language, which researchers must consider when analyzing oral data.

Additionally, as noted in Table 3, Lexical Complexity was examined using online tools based on the JACET 8000 scale (Uemura & Ishikawa, 2004). Developed by the Japan Association of College English Teachers (JACET), the JACET 8000 is a vocabulary list divided into eight levels according to word frequency across multiple corpora. Each level contains 1,000 words, ranging from the most common (Level 1) to the least frequent (Level 8). By analyzing the proportion of words students use in each category, researchers can estimate both the breadth and depth of their vocabulary. Lastly, a type-token ratio (TTR) indicates the richness of vocabulary by comparing the number of unique words (types) to the total number of words (tokens).

Results

Differences in Fluency from the Pre-Test to the Post-Test

As shown in Tables 4 and 5, the group that received TBLT significantly improved their Fluency across the board. Regarding Temporal Fluency (shown in Table 4), from the Pre-Test to the Post-Test, the TBLT group's mean length of runs increased by 1.28 syllables, and their speech

rate increased by 1.12 syllables per second. Noticeably, concerning this group's mean length of runs, the standard deviation for the Pre-Test was 1.13, indicating relatively low variability in scores, while the Post-Test had a standard deviation of 3.41, suggesting greater variability in the responses.

Table 4
Fluency (Temporal) from the Pre-Test to the Post-Test of the TBLT Group

N = 33 Category	Pre-Test Mean (SD)	Post-Test Mean (SD)	Mean Diff.	(%)	t	p	d
Mean length of runs	4.4 (1.13)	5.68 (3.41)	1.28	10.24	-2.641	.013	-.460
Speech rate	2.01 (.46)	3.13 (1.1)	1.12	35.78	-5.429	<.001	-.945

Further, concerning Hesitation Fluency (shown in Table 5), the TBLT group, on average, produced 6.43 fewer hesitation devices from the Pre-Test to the Post-Test. The standard deviation (SD) decreased from the Pre-Test (SD = 8.39) to the Post-Test (SD = 5.43), indicating less variability in scores after the intervention. Since the results were statistically significant and the effect sizes were medium to large, the observed differences are likely meaningful.

Table 5
Fluency (Hesitation) from the Pre-Test to the Post-Test of the TBLT Group

N = 33 Category	Pre-Test Mean (SD)	Post-Test Mean (SD)	Mean Diff.	(%)	t	p	d
Repetitions	6.76 (5.31)	4.15 (3.12)	2.61	38.61	4.224	.001	-.735
False starts	.3 (4.7)	.61 (.7)	.31	50.82	-.33	.067	-.576
Reformulations	.67 (.89)	.82 (.95)	.15	18.29	-.657	.516	-.114
Replacements	2.12 (2.16)	1.76 (1.48)	.36	16.98	.953	.348	.166
Unfilled pauses	3.76 (1.77)	2.3 (1.24)	1.46	38.83	3.689	<.001	.642
Filled pauses	4.97 (1.21)	2.52 (1.09)	2.45	49.2	9.521	<.001	1.657
Overall	18.58 (8.39)	12.15 (5.43)	6.43	34.61	5.494	<.001	.956

As reported in Tables 6 and 7, the group that received PPP instruction did not show significant improvement in any of the Fluency categories. Regarding Temporal Fluency (shown in Table 6), from the Pre-Test to the Post-Test, the PPP group's mean length of runs increased by only .006 syllables, and their speech rate increased by .024 syllables per second. For both of these categories, both the Pre-Test and Post-Test had low standard deviations, indicating that the participants' scores were tightly clustered around the mean, with little variability in their responses.

Moreover, concerning Hesitation Fluency (shown in Table 7), the PPP group, on average, produced .15 fewer hesitation devices from the Pre-Test to the Post-Test. The Pre-Test had a standard deviation of 9.59, indicating relatively high variability in scores, while the Post-Test had a standard deviation of 6.08, suggesting a reduction in variability and more consistency in the participants' responses. Since the results were not statistically significant and the effect sizes were small, the observed differences are likely negligible.

Table 6
Fluency (Temporal) from the Pre-Test to the Post-Test of the PPP Group

N = 33 Category	Pre-Test Mean (SD)	Post-Test Mean (SD)	Mean Diff.	(%)	t	p	d
Mean length of runs	3.279 (.996)	3.285 (.811)	.006	.18	-.034	.973	-.006
Speech rate	1.823 (.48)	1.847 (.433)	.024	1.3	-.364	.718	-.063

Table 7
Fluency (Hesitation) from the Pre-Test to the Post-Test of the PPP Group

N = 33 Category	Pre-Test Mean (SD)	Post-Test Mean (SD)	Mean Diff.	(%)	t	p	d
Repetitions	7.73 (5.21)	7.67 (3.61)	.06	.78	.097	.923	.017
False starts	.636 (.742)	.606 (.659)	.03	4.72	.205	.839	.036
Reformulations	1.33 (1.78)	.97 (.883)	.36	27.07	1.557	.129	.271
Replacements	2.15 (2.2)	2.18 (1.38)	.03	1.38	-.107	.916	-.019
Unfilled pauses	5.06 (2.32)	4.67 (2.68)	.39	7.71	.856	.339	.149
Filled pauses	4.52 (3.49)	5.18 (2.58)	.66	12.74	-.99	.33	-.172
Overall	21.42 (9.59)	21.27 (6.08)	.15	.7	.148	.883	.026

Differences in Syntactic Complexity from the Pre-Test to the Post-Test

As reported in Table 8, in terms of Syntactic Complexity from the Pre-Test to the Post-Test, the TBLT group, on average, increased their number of words per AS-Unit by 1.89, and they, on average, increased their number of subordinate clauses per AS-Unit by .115. Both of these increases were statistically significant, and effect sizes were large and medium, respectively. This indicates that the observed changes were both reliable and meaningful. The data sets in the categories presented in Table 8 all had low standard deviations, indicating minimal variability.

Table 8
Syntactic Complexity from the Pre-Test to the Post-Test of the TBLT Group

N = 33 Category	Pre-Test Mean (SD)	Post-Test Mean (SD)	Mean Diff.	(%)	t	p	d
Words/ AS-unit	4.44 (1.78)	6.33 (1.22)	1.89	29.86	-4.542	<.001	-.791
Subordinate clauses/ AS-unit	.115 (.106)	.23 (.196)	.115	50	-2.809	.008	-.489

As presented in Table 9, the data concerning the Syntactic Complexity of the PPP group showed mixed results. On the one hand, from the Pre-Test to the Post-Test, the PPP group, on average, increased their number of words per AS-Unit by .524, which was statistically significant (p = .004) and had a medium effect size (d = -.536). However, the average number of subordinate clauses per AS-Unit (.041) stayed the same. The data sets in the categories shown in Table 9 all exhibited low standard deviations, suggesting minimal variability.

Table 9
Syntactic Complexity from the Pre-Test to the Post-Test of the PPP Group

N = 33 Category	Pre-Test Mean (SD)	Post-Test Mean (SD)	Mean Diff.	(%)	t	p	d
Words/ AS-unit	4.034 (.824)	4.558 (.697)	.524	11.5	-3.081	.004	-.536
Subordinate clauses/ AS-unit	.041 (.085)	.041 (.117)	0	0	-.033	.974	-.006

Differences in Accuracy from the Pre-Test to the Post-Test

Table 10 presents the results concerning the TBLT group's Accuracy from the Pre-Test to the Post-Test. The TBLT group exhibited significant improvements in all three sub-categories of Accuracy. That is, the TBLT group, on average, increased their percentage of error-free clauses (by 15.49), decreased their number of errors per AS-Unit (by .166), and decreased their number of errors per 100 words (by 2.072). Regarding the Percentage of Error-Free Clauses category (in both the Pre-Test and Post-Test), the high standard deviations demonstrate considerable variability within the performances of this category. The differences in means across all three categories were statistically significant, with medium to large effect sizes, indicating reliable and meaningful improvements in Accuracy overall.

Table 10
Accuracy from the Pre-Test to the Post-Test of the TBLT Group

N = 33 Category	Pre-Test Mean (SD)	Post-Test Mean (SD)	Mean Diff.	(%)	t	p	d
Percentage of error-free clauses	64.22 (19.67)	79.71 (12.15)	15.49	19.43	-3.722	<.001	-.648
Errors/ AS-unit	.385 (.259)	.219 (.148)	.166	43.12	3.278	.003	.571
Errors/ 100 words	5.544 (3.339)	3.472 (1.761)	2.072	37.37	3.612	.001	.629

As reported in Table 11, similar to the TBLT group, the PPP group displayed significant improvements in all three sub-categories of Accuracy from the Pre-Test to the Post-Test. That is, the PPP group, on average, increased their percentage of error-free clauses (by 15.26), decreased their number of errors per AS-Unit (by 1.26), and decreased their number of errors per 100 words (by 1.47). Concerning the Percentage of Error-Free Clauses category (in both the Pre-Test and Post-Test), the high standard deviations demonstrate considerable variability within the performances of this category. As with the TBLT group, the differences in means across all three categories of the PPP group were statistically significant, with medium effect sizes, indicating reliable and meaningful improvements in Accuracy overall.

Table 11
Accuracy from the Pre-Test to the Post-Test of the PPP Group

N = 33 Category	Pre-Test Mean (SD)	Post-Test Mean (SD)	Mean Diff.	(%)	t	p	d
Percentage of error-free clauses	63.56 (11.79)	75.003 (9.56)	11.44	15.26	-4.946	<.001	-.861
Errors/ AS-unit	.395 (.162)	.269 (.095)	.126	31.9	4.04	<.001	.703
Errors/ 100 words	5.432 (2.33)	3.965 (1.81)	1.47	27	4.013	<.001	.699

Differences in Lexical Complexity from the Pre-Test to the Post-Test

Overall, both groups have shown some improvement in their Lexical Complexity. As Table 12 reports, from the Pre-Test to the Post-Test, the TBLT group uttered 11.06% fewer words in Level 1, 3.24% more words in Level 2, 1.91% more words in Level 3, .869% more words in Level 4, .373% more words in Level 5, .133% more words in Level 6, .182% more words in Level 7, and .036% more words in Level 8. All standard deviations were low, except for Level 1, which exhibited greater variability than the others.

Table 12
Lexical Complexity from the Pre-Test to the Post-Test of the TBLT Group

N = 33 Category	Pre-Test Mean (SD)	Post-Test Mean (SD)	Mean Diff.	(%)	t	p	d
JACET 8000 Level 1	84.31 (6.35)	73.25 (11.53)	11.06	13.18	5.829	<.001	1.015
JACET 8000 Level 2	5.235 (2.92)	8.475 (3.33)	3.24	38.23	-5.014	<.001	-.873
JACET 8000 Level 3	.555 (.72)	2.461 (1.54)	1.91	77.46	-7.213	<.001	-1.256
JACET 8000 Level 4	.566 (.79)	1.435 (1.988)	.869	60.56	-2.202	.035	-.383
JACET 8000 Level 5	.065 (.273)	.438 (.911)	.373	85.11	-2.186	.036	-.381
JACET 8000 Level 6	.090 (.249)	.223 (.406)	.133	59.64	-1.473	.151	-.256
JACET 8000 Level 7	0 (0)	.182 (.45)	.182	100	_	_	_
JACET 8000 Level 8	0 (0)	.036 (.145)	.036	100	_	_	_
Out of the JACET list	7.95 (6.93)	3.21 (3.13)	4.74	59.62	3.806	<.001	.663
Type-Token Ratio (TTR)	.343 (.184)	.579 (.107)	.236	40.74	-7.563	<.001	-1.317

As the TBLT group used fewer Level 1 (easier/more frequent) words and more (difficult/less frequent) words in the subsequent Levels 2-8 in the Post-Test than in the Pre-Test, this is an indicator that this group had improved their Lexical Complexity. The mean differences from Level 1 to Level 4 were all statistically significant. Levels 1 to 3 showed large effect sizes,

while Level 4 showed a small to medium effect size. These results suggest substantial and meaningful improvements across the levels. Similarly, concerning the Type-Token Ratio (which indicates the degree to which participants used different words), the TBLT group improved (by .236) from .343 in the Pre-Test to .579 in the Post-Test. This difference in means was statistically significant, with a large effect size.

As shown in Table 13, the PPP group also showed considerable improvement where Lexical Complexity was concerned. For instance, from the Pre-Test to the Post-Test, the PPP group uttered 2.72% fewer words in Level 1, 1.64% more words in Level 2, .417% more words in Level 3, .011% more words in Level 4, .259% more words in Level 5, .028% more words in Level 6, .008% more words in Level 7, and .001% more words in Level 8. All standard deviations were low overall, indicating limited variability. Similar to the TBLT group, though with slightly smaller gains, the PPP group used fewer Level 1 (easier/more frequent) words and increased their use of more (difficult/less frequent) words in Levels 2–8 in the Post-Test compared to the Pre-Test. The mean differences for Level 1 to Level 3 were all statistically significant. The effect sizes for Levels 1 and 2 were medium, while the effect size for Level 3 was small, suggesting that the observed changes were meaningful, though with a diminishing impact at higher levels. Lastly, regarding the Type-Token Ratio, the PPP group improved (by .084) from .387 in the Pre-Test to .471 in the Post-Test. This mean difference was statistically significant, demonstrating a large effect size.

Table 13
Lexical Complexity from the Pre-Test to the Post-Test of the PPP Group

N = 33 Category	Pre-Test Mean (SD)	Post-Test Mean (SD)	Mean Diff.	(%)	t	p	d
JACET 8000 Level 1	83.79 (2.51)	81.07 (2.81)	2.72	3.35	4.079	<.001	.71
JACET 8000 Level 2	6.58 (1.63)	8.22 (2.18)	1.64	19.9	-3.665	<.001	-.638
JACET 8000 Level 3	.846 (.642)	1.263 (.868)	.417	33	-2.436	.021	-.424
JACET 8000 Level 4	.594 (.705)	.605 (.7)	.011	1.82	-.068	.946	-.012
JACET 8000 Level 5	.066 (.286)	.04 (.181)	.259	39.39	.427	.672	.074
JACET 8000 Level 6	.137 (.433)	.141 (.443)	.028	2.84	-.035	.972	-.006
JACET 8000 Level 7	.125 (.384)	.117 (.294)	.008	6.4	.085	.933	.015
JACET 8000 Level 8	.082 (.265)	.083 (.240)	.001	121	-.018	.985	-.003
Out of the JACET list	7.79 (2.9)	8.47 (2.18)	.68	8.03	-1.065	.295	-.185
Type-Token Ratio (TTR)	.387 (.127)	.471 (.132)	.084	17.83	-3.139	.004	-.546

Summary and Discussion

In summarizing and interpreting the findings of this current study, the research question will be answered and discussed.

How do TBLT and PPP compare in affecting Japanese EFL university students' speaking skills in terms of Fluency (both temporal and hesitation), Syntactic Complexity, Accuracy, and Lexical Complexity in this study?

In short, this study found that both TBLT and PPP had a positive effect on various aspects of learners' EFL oral proficiency. However, overall, the TBLT outperformed the PPP group. The following will outline and discuss the differences between the two groups in each of the sub-categories presented in the research question.

First, concerning Fluency, as we hypothesized, this is an area where the TBLT group showed significant improvement, but the PPP group did not. Regarding Temporal Fluency, as Figure 1 illustrates, the TBLT significantly increased their mean length of runs and speech rate while the PPP did not. Similarly, as the overall Hesitation devices shown in Figure 2 demonstrate, the TBLT group significantly reduced the number of their Hesitation devices, while the PPP group did not. These results were not surprising, as the strength of TBLT has long been thought to be its emphasis on the use of authentic language through meaningful tasks to promote fluency and student confidence (Gonz âlez-Lloret and Nielson, 2015).

The data regarding Syntactic Complexity, as shown in Figure 3, also shows a clear advantage of TBLT over the PPP method. That is, the TBLT group was able to increase their number of words per AS-Unit significantly, while the PPP group only registered a moderate increase. Further, while the overall numbers in this category were small, the TBLT group significantly increased their number of subordinate clauses per AS-Unit, while the PPP group did not.

Figure 1
Comparing the Temporal Fluency of TBLT and PPP Groups Over Time

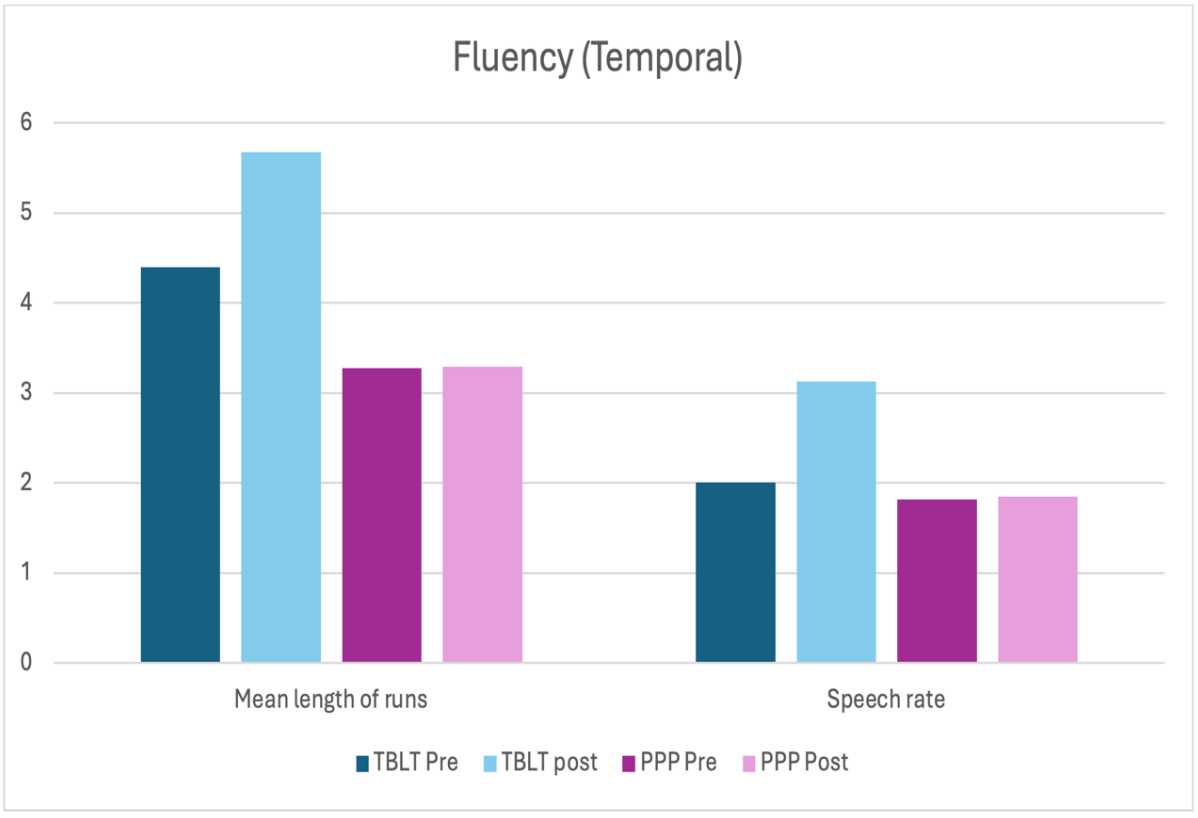


Figure 2
Comparing the Hesitation Fluency of TBLT and PPP Groups Over Time

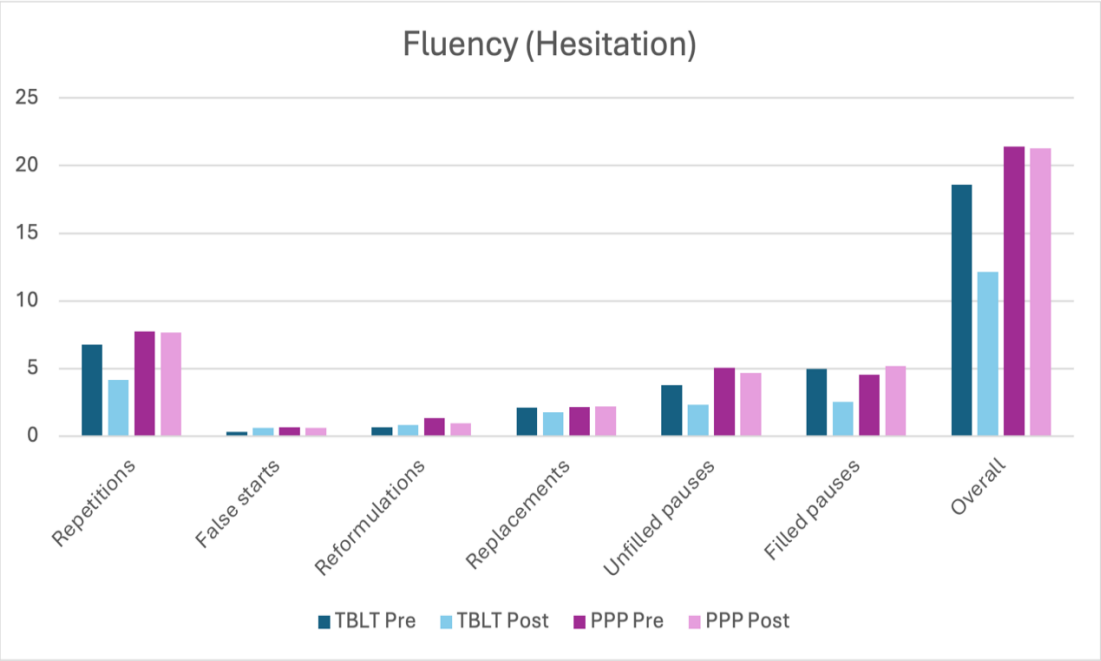
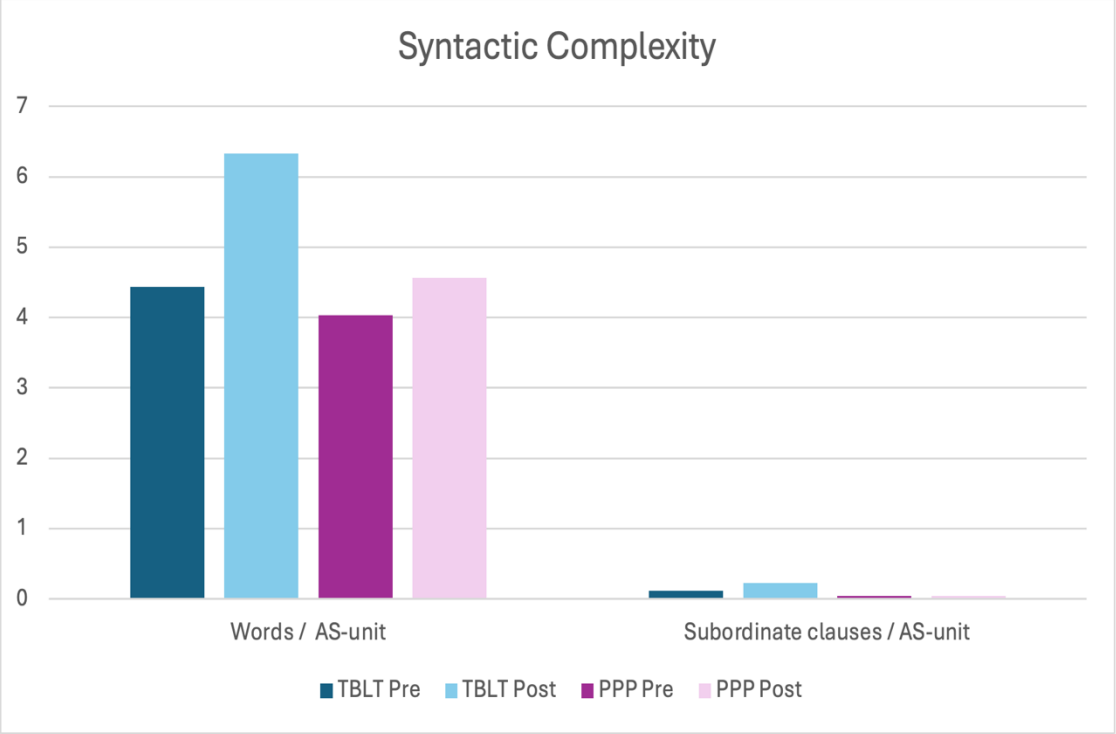


Figure 3
Comparing the Syntactic Fluency of TBLT and PPP Groups Over Time



Similar to Syntactic Complexity, there might be an expectation of PPP outperforming TBLT in Accuracy due to PPP's emphasis on explicit instruction and controlled practice, which are designed to reinforce correct language use and minimize errors. Conversely, TBLT's focus on communication has been thought to lead to greater fluency but at the potential cost of accuracy. However, the findings of this study did not support this assumption with regard to Accuracy. As illustrated in Figures 4, 5, and 6, both groups demonstrated significant gains in all three

Accuracy measurements (% of error-free clauses, errors per AS-Unit, and errors per 100 words) in this study. In all three categories, the TBLT group slightly outperformed the PPP group.

Figure 4
Comparing Accuracy (% of Error-Free Clauses) of TBLT and PPP Groups Over Time

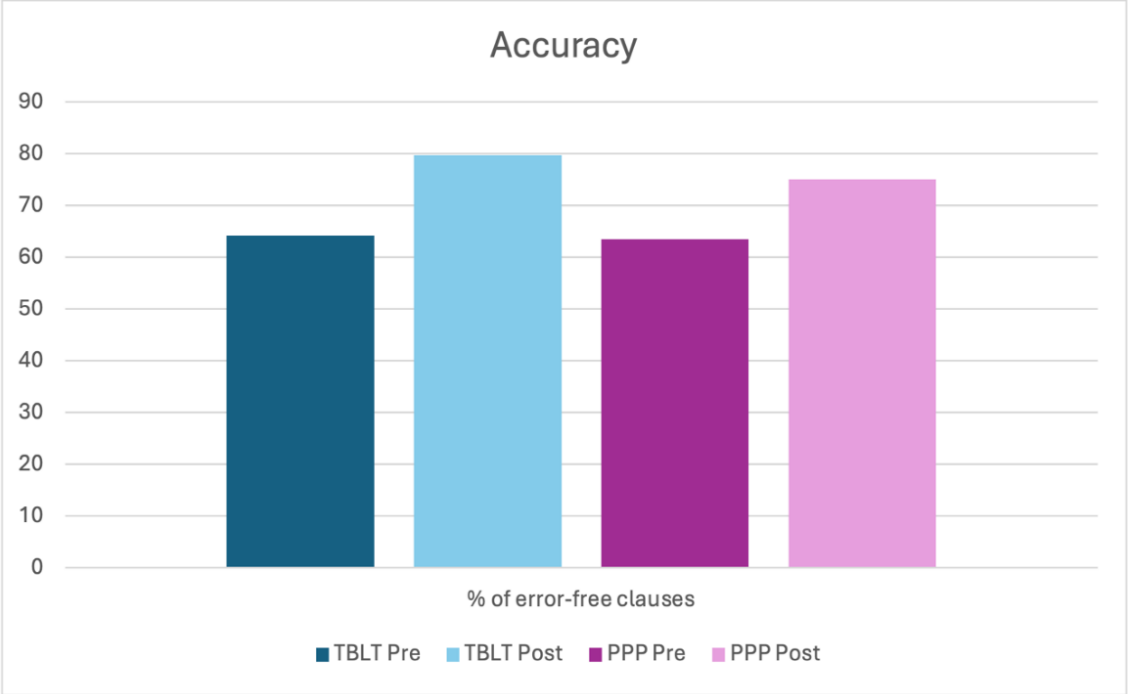
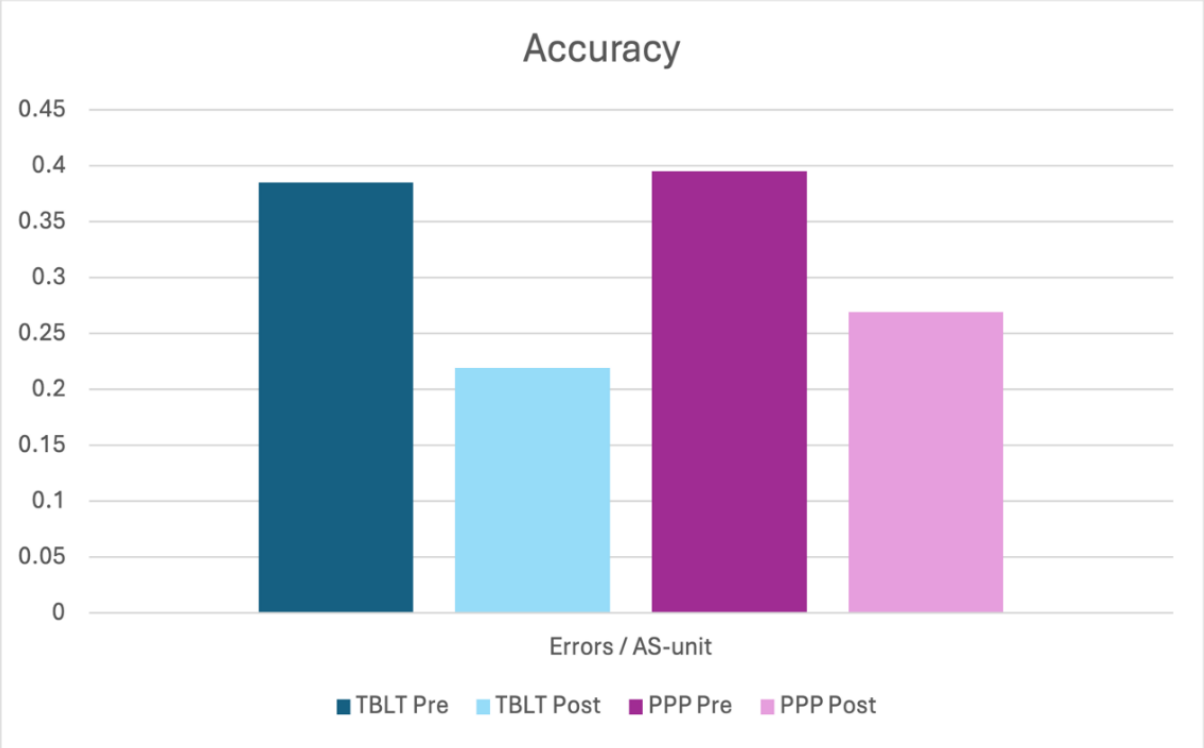
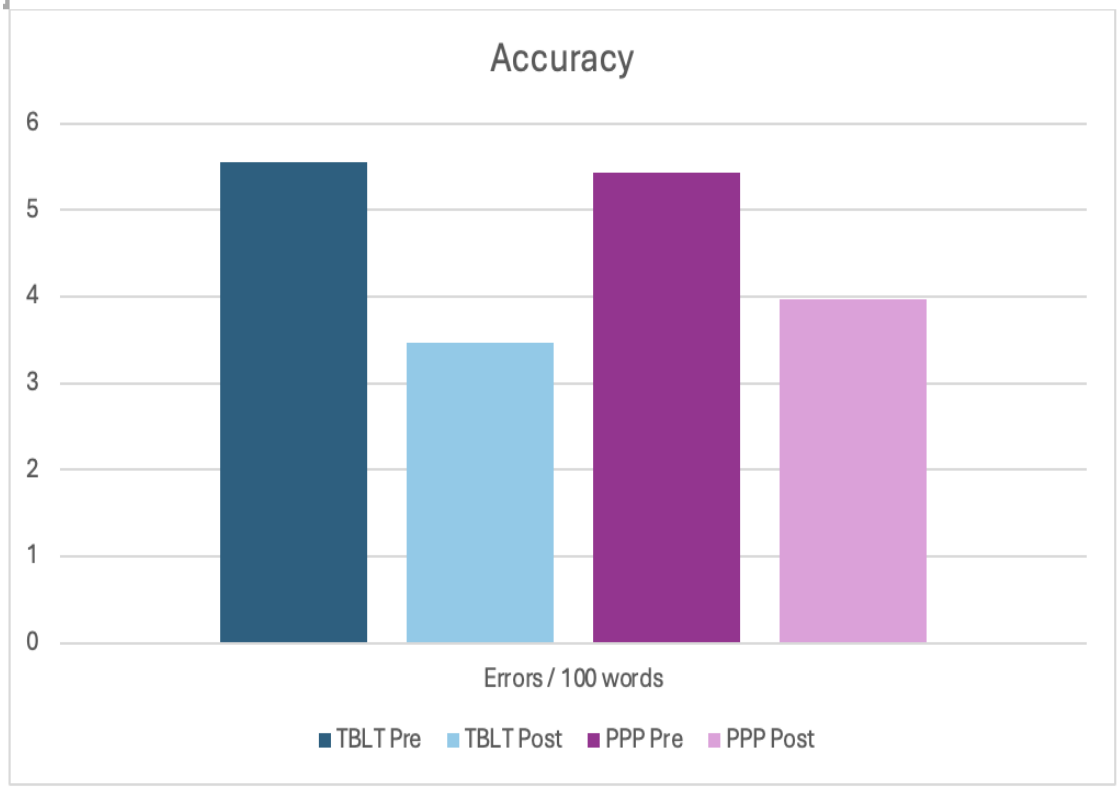


Figure 5
Comparing Accuracy (Errors per AS-Unit) of TBLT and PPP Groups Over Time



 ISSN 2790-9441

Figure 6
Comparing Accuracy (Errors per 100 Words) of TBLT and PPP Groups Over Time



Thus, TBLT's advantage over PPP in both Syntactic Complexity and Accuracy is somewhat surprising since PPP's explicit instruction and controlled practice are expected to support more complex and accurate language use. In contrast, TBLT's focus on communication rather than direct grammatical instruction would suggest lower gains in both areas, challenging the assumption that structured practice is essential for developing complex and accurate grammar. These findings align with studies by Arslanyilmaz (2013) and González-Lloret and Nielson (2015), which also demonstrated the benefits of TBLT in enhancing learners' Syntactic Complexity and Accuracy.

Similar to most of the categories presented above, the findings regarding Lexical Complexity suggest an overall advantage towards TBLT. As shown in Figure 7, although both groups improved by using more common and simpler words in the Pre-Test and fewer common, more complex words in the Post-Test, the TBLT group showed significantly greater gains and outpaced the PPP group in this sub-category overall. Likewise, as shown in Figure 8, both groups made notable improvements in TTR over time, but the TBLT group ultimately achieved greater gains.

Consistent with the results of Shintani (2013, 2015), the findings in this study provide further evidence supporting the idea that TBLT may be more effective than PPP for vocabulary acquisition. Shintani (2015) explains this by using cognitive load theory, specifically the *need*, *search*, and *evaluate* framework. In the PPP lessons, learners *need* to use the target items but do not *search* for their meanings since the teacher provided them. In contrast, the TBLT lessons require learners not only to *need* the items but also to actively *search* for their meanings. Both approaches involve *evaluation*, but PPP limits this process because it depends on teacher feedback. In TBLT, learners have to engage more deeply in *evaluation*, such as when they try

to infer the meaning of an adjective after failing to understand a command, and when they use adjectives to clarify commands during communication breakdowns.

Figure 7
Comparing Lexical Complexity (JACET 8000) of TBLT and PPP Groups Over Time

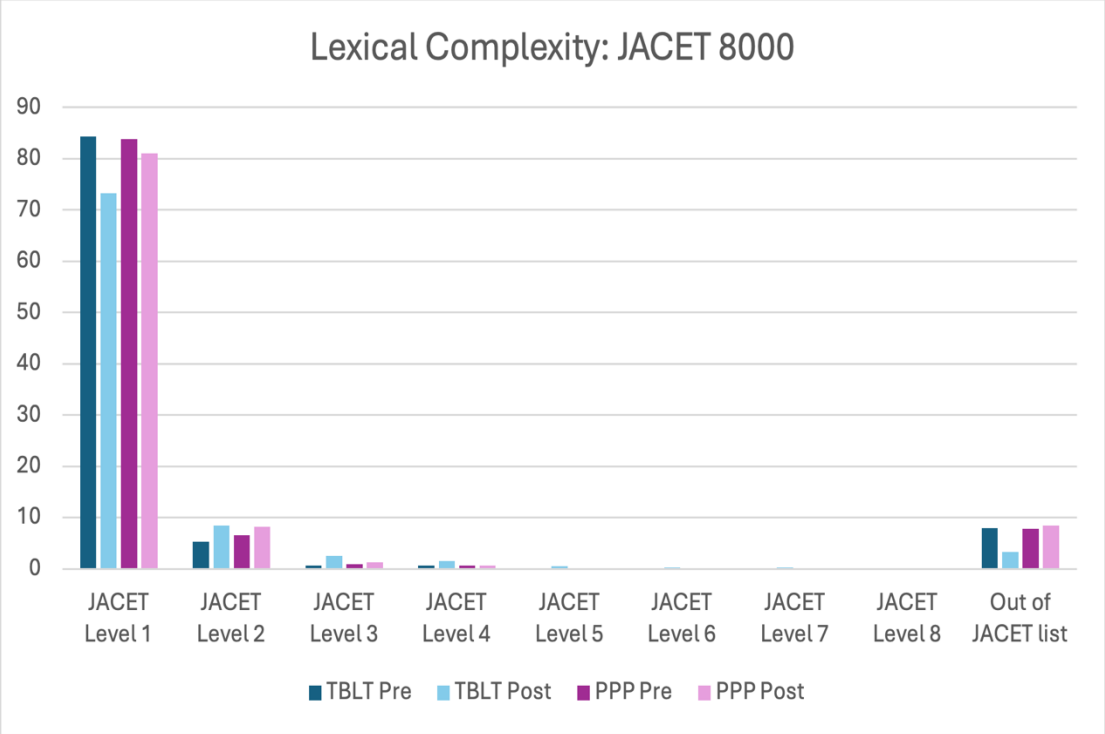
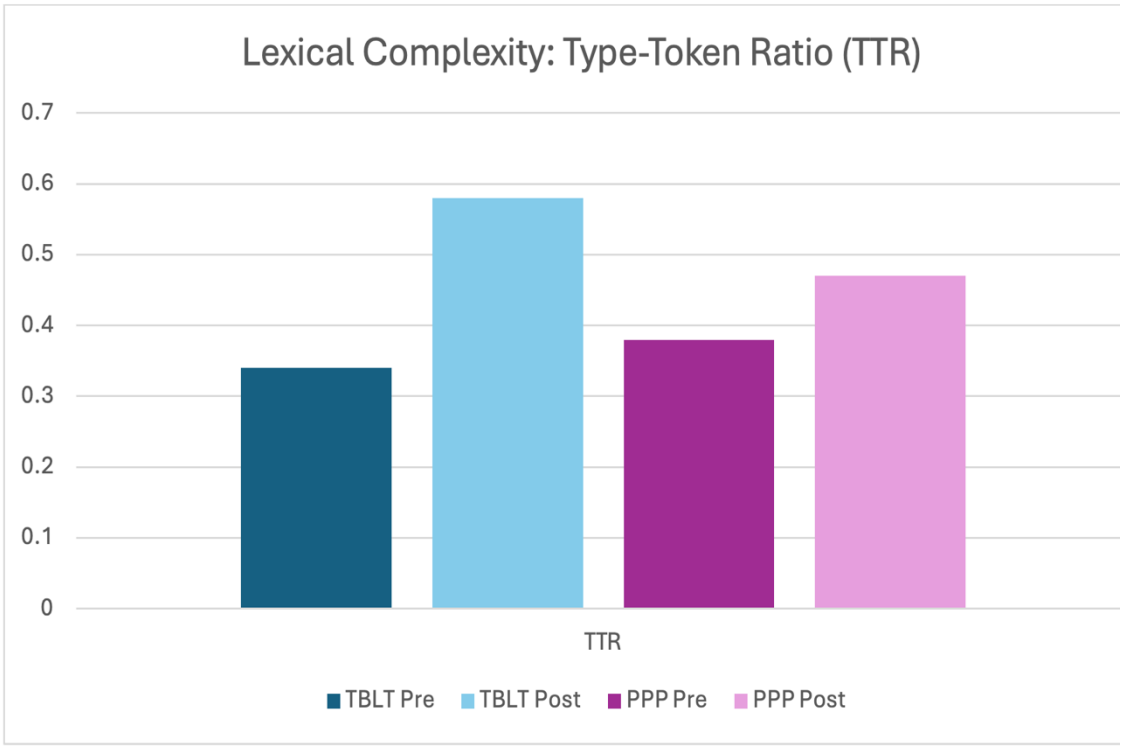


Figure 8
Comparing Lexical Complexity (Type-Token Ratio) of TBLT and PPP Groups Over Time



 ISSN 2790-9441

Conclusion and Implications

In conclusion, this study highlights the vast potential of TBLT to enhance oral proficiency among Japanese EFL university students. By offering additional practical evidence supporting TBLT, it reinforces the psycholinguistic foundation for TBLT and underscores its real-world benefits for EFL learners in Japan. Based on these findings, the writer encourages language instructors in Japan to consider incorporating TBLT into their English communication classes.

Nonetheless, the authors acknowledge that implementing TBLT presents several challenges. Despite its substantial history and research backing, TBLT remains poorly understood in Japan, with persistent misconceptions (Ellis, 2014, 2021a). Additionally, EFL instructors and material designers may find it difficult to shift away from the familiar PPP approach. As Ellis (2021a) observes, the availability of genuine TBLT coursebooks in the Japanese EFL market is limited. Therefore, a practical first step for educators would be to enhance their understanding of TBLT and gradually integrate its elements into their teaching.

Accordingly, to accommodate PPP traditionalists who may be resistant to change, a gradual and/or mixed approach may be one way forward. Ellis (2019, 2021b) makes a strong case for a modular curriculum consisting of separate (i.e., non-integrated) task-based and structure-based components. This approach emphasizes developing fluency first through a task-based module that includes a focus on form (Long, 1991). Once learners have reached a basic level of fluency, a second, structure-based module is added to provide explicit accuracy-oriented work to counteract learned selective attention (N. Ellis, 2006).

While this study provides valuable insights, further research is needed to confirm and expand upon these findings. For instance, this study's small sample size, limited to first-year students from a single faculty at one university, restricts the generalizability of the results. Future research should explore larger and more diverse student populations across various faculties and institutions. Additionally, longitudinal studies examining a wider array of target features of language would be useful. Lastly, the absence of a control group and a delayed post-test due to logistical constraints limits the strength of causal inferences and the assessment of long-term effects. Including these elements in future studies would enhance the reliability and depth of the findings. The authors hope that studies like this one will inspire further investigation and encourage broader adoption of TBLT in the Japanese EFL context.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Number 22K00685.

References

Andersen, J. (2016). A potted history of PPP with the help of ELT Journal. *ELT Journal, 71*(2), 218-227.

Arslanyilmaz, A. (2013). Computer-assisted foreign language instruction: Task-based vs. form focused. *Journal of Computer Assisted Learning, 29*(1), 303–18.

Beretta, A., & Davies, A. (1985). Evaluation of the Bangalore project. *ELT Journal, 39*(1), 121-127.

Brown, G., Anderson, A., Shilcock, R., & Yule, G. (1984). *Teaching talk: Strategies for production and assessment*. Cambridge University Press.

Bygate, M. (1987). *Speaking*. Oxford University Press.

Byrne, D. (1976). *Teaching oral English*. Longman.

Caine, N. A. (2005). *EFL examination washback in Japan: Investigating the effects of oral assessment on teaching and learning*. Unpublished Master's Dissertation, University of Manchester, Manchester.

Case, A. (2008). 15 variations of PPP. *Tefl.NET*, *15*(1). Available from: https://www.tefl.net/elt/ideas/teaching/variations-on-ppp/ (Accessed March 2025)

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Cutrone, P. (2018). Insights into the field of applied linguistics: An interview with Rod Ellis. *Journal of the School of Global Humanities and Social Sciences, Nagasaki University, 4*(1) 309-319.

Cutrone, P., & Beh, S. (2015). *Welcome to Kyushu, Japan: A task-based approach to EFL learning using authentic dialogues*. Shohakusha.

Cutrone, P., & Beh, S. (2017). Adopting a task-based approach to address ELT issues in Japan. In Leung, Y. (Ed.), *Selected Papers from the Twenty-sixth International Symposium on English Teaching* (pp. 150-159). English Teachers' Association – Republic of China (ETA_ROC): Crane Publishing.

Cutrone, P., & Beh, S. (2018). Investigating the effects of task-based language teaching on Japanese EFL learners' willingness to communicate. *The Journal of Asia TEFL, 15*(3), 566-589.

Cutrone, P., & Beh, S. (2024). The effects of task-based instruction on Japanese EFL learners' communicative confidence and willingness to communicate. *International Journal of TESOL Studies, 6*(1), 15-33.

De la Fuente, M. J. (2006). Classroom L2 vocabulary acquisition: Investigating the role of pedagogical tasks and form-focused instruction. *Language Teaching Research 10*(1), 263–95.

De Ridder, I., Vangehuchten, L., & Gomez, M. (2007). Enhancing automaticity through task-based language learning. *Applied Linguistics, 28*(1), 309–15.

Ellis, N. (2006). Selective attention and transfer phenomena in L2 acquisition: contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning. *Applied Linguistics, 27*(2), 164-194

Ellis, R. (1987). Interlanguage variability in narrative discourse: Style shifting in the use of the past tense. *Studies in Second Language Acquisition, 9*(1), pp. 12-20.

Ellis, R. (1991). Communicative competence and the Japanese learner. *JALT Journal,13*(1), 103-127.

Ellis, R. (2006). The methodology of task-based teaching. *Asian EFL Journal, 8*(3), 19-45.

Ellis, R. (2009). Task-based language teaching: Sorting out the misunderstandings. *International Journal of Applied Linguistics, 19*(3), 221-246.

Ellis, R. (2014) Taking the critics to task: The case for task-based teaching. *Proceedings of CLaSIC 2014*, 103-117.

Ellis, R. (2018). *Reflections on task-based language teaching (second language acquisition, 125).* Bristol: Multilingual Matters.

Ellis, R. (2019). Towards a modular language curriculum for using tasks. *Language Teaching Research, 23*(4), 454-475.

Ellis, R. (2021a). Task-based language teaching: Early days, now and into the future. In N. P. Sudharshana, & L. Mukhopadhyay (Eds.), *Task-based language teaching and assessment: Contemporary reflections from across the world* (pp. 39-61). Singapore: Springer.

Ellis, R. (2021b). Options in a task-based language-teaching curriculum: An educational perspective. *TASK: Journal of Task-Based Language Teaching and Learning, 1*(1), 11-46.

Ellis, R., Skehan, P., Li, S., Shintani, N., and Lambert, C. (2020). *Task-based language teaching: Theory and practice*. Cambridge: Cambridge University Press.

Farooq, M. U. (2005). A model for motivating Japanese EFL learners through real-life questioning strategies. *The Journal of Nagoya Gakkugei University, 1*(1), 27-42.

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics, 21*(1), 354-375.

Fries, C. (1945). *Teaching and learning English as a foreign language*. Michigan: University of Michigan Press.

Fries, C. (1952). *The structure of English*. New York: Harcourt Brace & Company.

González-Lloret, M., & Nielson, K. (2015). Evaluating TBLT: The case of a task-based Spanish program. *Language Teaching Research 19*(1), 525–49.

Harris, B. (2015). *'Where are we now? Current teaching paradigms in pre-service training'*. Paper presented at the 49th International IATEFL Annual Conference, Manchester, UK.

Harris, J., & Leeming, P. (2021). The impact of teaching approach on growth in L2 proficiency and self-efficacy: A longitudinal classroom-based study of TBLT and PPP. *Journal of Second Language Studies, 5*(10), 1-30.

Helgesen, M., Brown, S., & Mandeville, T. (2007). *English firsthand*. Tokyo: Pearson Longman.

Inoue, C. (2010). Investigating the sensitivity of the measures of fluency, accuracy, complexity and idea units with a narrative task**.** *Papers from the Lancaster University Postgraduate Conference in Linguistics & Language Teaching, 4*(1), 1-24.

JASP Team (2024). JASP (Version 0.18.3)[Computer software].

Long, M. (1985). Input and second language acquisition theory. In S. Gass, & C. Madden (Eds.). *Input in second language acquisition*. Newbury House. pp. 377–393.

Long, M. (1991). Focus on form: A design feature in language teaching methodology. In K. De Bot, R. Ginsberg, & C. Kramsch (Eds.), Foreign language research in cross-cultural perspectives (pp. 39-52). John Benjamins.

Long, M. H., Lu, S., & Yi, W. (2017). Questions and answers on task-based language teaching. *International Chinese Language Education, 2*(1), 88-99.

Martin, D. (2003). *Talk a lot*. EFL Press.

Mizuho, A. (2017). Japan's latest English-proficiency scores disappoint. *The Japan Times*. Available from: https://www.japantimes.co.jp/news/2017/04/06/national/japans-latest-english-proficiency-scores-disappoint/ (Accessed January 2024)

Mudinillah, A., Rahmi, S., & Taro, N. (2024). Task-based language teaching: A systematic review of research and applications. Lingeduca: *Journal of Language and Education Studies, 3*(1), 102-115.

Nunan, D. (1988). *Syllabus design*. Oxford University Press.

Nunan, D. (2004). *Task-based language teaching*. Cambridge University Press.

Pica, T. Kanagy R., & Falodun, J. (1993). Choosing and using communicative tasks for second language acquisition. In S. Gass, & G. Crookes (Eds.), *Tasks and language learning: Integrating theory and practice* (pp. 9-34). Clevedon: Multilingual Matters.

Richards, J. C., & Rodgers, T. S. (2014). *Approaches and methods in language teaching* (3rd ed.). Cambridge University Press.

Reesor, M. (2002). The bear and the honeycomb: A history of Japanese English language policy. *NUCB JLCC (Nagoya University of Commerce and Business Journal of Language, Culture and Communication), 4*(1), 41-52.

Robinson, P. (1998). State of the Art: SLA theory and second language syllabus design. *The Language Teacher, 22*(4), 7-14.

Sakui, K. (2004). Wearing two pairs of shoes: language teaching in Japan. *ELT J, 58*(2), 155-163.

Sato, R. (2010). Reconsidering the effectiveness and suitability of PPP and TBLT in the Japanese EFL classroom. *JALT Journal*, *32*(2), 189-200.

Shintani, N. (2013). The effect of focus on form and focus on forms instruction on the acquisition of productive knowledge of L2 vocabulary by young beginning level learners. *TESOL Quarterly 47*(1), 36–62.

Shintani, N. (2015). The incidental grammar acquisition in focus on form and focus on forms instruction for young beginner Learners. *TESOL Quarterly 49*(1), 115–40.

Sheen, R. (2005). Focus on forms as a means of improving accurate oral production. In Housen, A., & Pierrard, M. (Eds.), *Investigations in instructed language acquisition* (pp. 270–310). Mouton de Gruyter.

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford University Press.

Soars, J. & L. Soars (2016). *American headway 2 (3rd edition)*. OUP.

Swain, M. (1993). The output hypothesis: Just speaking and writing aren't enough. *Canadian Modern Language Review*, *50*(1): 158–164

Taylor, S. (2020). Factors affecting the speaking fluency of Japanese EFL learners. *Language Education and Research Center Journal*, *15*(1), 3-25.

Uemura, T., & S. Ishikawa (2004). JACET 8000 and Asia TEFL vocabulary initiative. *The Journal of Asia TEFL, 1*(1), 333-347.

Van den Branden, K. (2006). *Task-based language teaching: From theory to practice*. Cambridge: Cambridge University Press.

Willis, D. (1993). Syllabus, corpus, and data-driven learning. In Kennedy, C. (Ed.), *Plenaries from the 1993 IATEFL conference* (pp. 47-65). Swansea, Wales.

Willis, J. (1998), Task-based learning: What kind of adventure? *The Language Teacher*, *22*(7), 17-18.

Willis, J., & Willis, D. (2007). *Doing task-based teaching.* Oxford University Press.

Pino Cutrone is an Associate Professor in the School of Global Humanities and Social Sciences at Nagasaki University. His research interests include pragmatics, discourse analysis, intercultural communication, sociolinguistics relating to Japanese EFL learners, educational technology, and study abroad programs. He has published widely in the field of applied linguistics and intercultural pragmatics.