

Articles

Temporal Dynamics of EFL Learners' Oral Proficiency: Effects of Coursework and Self-Training Using Shadowing-Based Tasks

Noriko Nakanishi
Kobe Gakuin University

In EFL (English as a Foreign Language) environments, opportunities for learners to use the language outside the classroom are often limited, making it challenging to foster oral proficiency. This study examined the effects of spring semester coursework and summer self-training on the oral proficiency of 217 Japanese university students. It addressed 2 research questions: (1) changes in oral proficiency across the spring semester and subsequent summer vacation, and (2) the effects of intensive self-training during the vacation period. Semester instruction followed a flipped learning model, and summer training consisted of shadowing and related oral tasks. Oral proficiency was assessed using the Versant Speaking Test sub-scores. Results showed that mean scores improved during the spring semester but either plateaued or declined during the summer break. However, learners who actively engaged in summer self-training demonstrated small gains in fluency and pronunciation, suggesting the potential of autonomous practice to help sustain procedural oral skills.

英語が外国語として学ばれる環境(EFL)では、教室外での使用機会が限られるため、学習者の口頭能力の発達を促すことが難しい。本研究では、日本の大学生217名を対象に、(1) 春学期と夏季休暇を通じた口頭能力の変化、(2) 休暇中の集中的な自己トレーニングの効果を検討した。学期中はflipped learning型の授業を行い、夏休みにはシャドーイングおよび関連課題による自己練習を行った。Versant Speaking Testの4つのサブスコアを4月、7月、9月に測定した結果、春

<https://doi.org/10.37546/JALTJJ47.2-1>

JALT Journal, Vol. 47, No. 2, November 2025

学期には全体的な向上が見られたが、休暇中は停滞または低下が生じた。一方、夏休み中の自己練習に積極的に取り組んだ学習者は、流暢さと発音において小さな向上を示した。これらの結果は、教室活動と自己練習の両立が口頭能力維持に有効であることを示唆している。

Keywords: self-training; semester transitions; shadowing; vacation impact; Versant Speaking Test

In language learning, the development of oral language proficiency, including fluency, pronunciation, and lexical knowledge, typically requires sustained exposure and repeated practice over time (Bygate, 2001; DeKeyser, 2007; Nation, 2007; Swain, 1985). According to Hakuta et al. (2000), ESL (English as a Second Language) learners in California and Canada needed three to five years to develop oral proficiency. If such extended timelines are required in ESL environments, it is evident that even more training time is necessary in EFL (English as a Foreign Language) contexts, where exposure to the target language is much more limited.

Given these challenges, the present study aims to examine changes in EFL learners' oral English proficiency over a six-month period by investigating the effects of interactive coursework during the spring semester and voluntary self-training during the summer vacation. The study focused on 217 Japanese university students, all of whom participated in classroom-based learning. During the semester, interactive classroom lessons were conducted based on a flipped learning approach (Lee & Wallace, 2018), in which students completed preparatory tasks such as listening comprehension, shadowing practice, and oral presentation rehearsals prior to class. For the self-training component, only students who opted in completed online assignments incorporating shadowing-based tasks (Minematsu et al., 2023). Learners' oral proficiency was assessed using the sub-scores of the Versant Speaking Test: Fluency, Pronunciation, Sentence Mastery, and Vocabulary (Pearson, 2022).

Literature Review

Limited Opportunities for Oral Communication in EFL Contexts

EFL learners typically encounter fewer opportunities for oral communication than their ESL counterparts. Lightbown and Spada (2013) emphasize that EFL learners receive only a few hundred hours of language instruction over several years, compared to the immersive environments found in ESL contexts. Furthermore, when both teachers and students share the same L1, as is often the case in Japan, classroom communication

frequently defaults to the native language, reducing learners' chances to engage in meaningful target language use (Eyring, 2014).

These environmental constraints have measurable consequences for learners' oral proficiency development. Zhang (2009) observed that limited access to authentic input and output in Chinese EFL settings hindered learners' development of spoken fluency. Similarly, Teeter (2017) found that Japanese university students experienced a decline in listening proficiency during their first year, due to structural limitations such as large class sizes, reading-focused instruction, and insufficient time devoted to speaking practice. These findings support the view that sustained and meaningful oral interaction is crucial for the development of fluency and accuracy (Bygate, 2001; Swain, 1985).

Muñoz (2014) further noted that it is not the age of exposure but rather the quantity and quality of interaction that predict oral performance. Her longitudinal research revealed that learners who experienced extended informal contact and time abroad achieved higher levels of fluency and syntactic complexity than those with classroom-only instruction. Such evidence supports Nation's (2007) assertion that fluency development requires a dedicated strand of practice-oriented input and output over time.

Structural limitations within academic calendars exacerbate this issue. For example, Japan's University Establishment Standards (MEXT, 1956, Chapter 6, Article 22) recommend a 35-week academic calendar per year, consequently leaving an approximately 17-week vacation period. Even if gains are made during the semester, students risk losing progress during long vacation periods without sustained input or practice opportunities. Cooper (2003) identified a similar phenomenon in the U.S.—“summer learning loss”—showing that extended breaks can lead to declines in academic and procedural skills, especially among learners from disadvantaged or non-English-speaking backgrounds. This concern is applicable to EFL learners as well: without structured input and practice during breaks, oral proficiency gains made during the semester may regress.

Effects of Self-Training

To mitigate learning loss during breaks, various studies have explored the effects of self-training on oral proficiency. Trofimovich et al. (2009) demonstrated that while long-term listening and reading comprehension support pronunciation development, production-based tasks, such as speaking, are essential for improving fluency and comprehensibility. Their

findings align with DeKeyser's (2007) claim that practice involving active output is necessary to automatize oral skills.

In a related study, Nagle et al. (2025) examined high-variability phonetic training (HVPT) and found that training frequency, rather than the number of talkers, had the greatest impact on perceptual learning outcomes. This reinforces the argument that both regularity and structure in self-training are key factors for success in EFL oral skill development.

Programs incorporating shadowing as a self-training technique have also shown promising results. Kuniyama et al. (2022) and Nakanishi et al. (2022) found that Japanese university students who actively participated in summer shadowing activities improved their oral fluency. Similarly, Foote and McDonough (2017) implemented mobile-assisted shadowing using iPods, resulting in noticeable gains in both fluency and comprehensibility, as well as positive learner attitudes.

The Role of Shadowing in Oral Proficiency Development

Shadowing has been defined as “a paced, auditory tracking task which involves the immediate vocalization of auditorily presented stimuli” (Lambert, 1992, p. 266) and is widely recognized as an effective method for enhancing oral proficiency. Hamada (2019a) reviewed both theoretical and empirical literature and emphasized that shadowing promotes deep cognitive processing by requiring focused attention on auditory input. This results in improvements in listening comprehension (Hamada, 2019b).

Kadota (2019) elaborated on four mechanisms by which shadowing facilitates language acquisition: input effects (enhanced listening), practice effects (improved phonological working memory), output effects (better fluency through speech simulation), and monitoring effects (metacognitive regulation of speech). These multifaceted effects position shadowing as a uniquely effective technique that bridges the gap between input and output, echoing Swain's (1985) comprehensible output hypothesis and Nation's (2007) emphasis on fluency-building activities.

By combining listening and speaking in an integrated task, shadowing supports both receptive and productive skills, making it particularly valuable in contexts where naturalistic exposure is limited. It also allows for flexibility in design, ranging from passive listening to active vocal production, offering educators a versatile tool to support oral proficiency development across various EFL contexts.

Research Questions

This study investigates how coursework during the spring semester and self-training during the summer vacation influence the oral proficiency of EFL learners. Two main research questions are addressed:

- RQ 1. How does oral proficiency change over the course of a semester and the following long vacation?
- RQ 2. Does intensive self-training during the vacation period affect oral proficiency?

The first research question focuses on the impact of in-class coursework and the summer training program. It is expected that while students' proficiency may improve during the semester, it might decline or stagnate during the long vacation period due to reduced opportunities for language use. Moreover, the study distinguishes between first- and second-year students in English conversation courses. Students in the first year, who take English conversation courses for the first time, experience a shift from traditional reading-heavy classes (Teeter, 2017) to more interactive speaking-focused lessons. This transition can impact their oral proficiency, as they are not only exposed to new learning methods but also to their first substantial exposure to real-time spoken English. Second-year students, who have had one year of experience in the English conversation course, may demonstrate different proficiency gains due to their previous exposure. Therefore, it is crucial to investigate how these two groups (first- and second-year students) respond differently to the semester's coursework, considering the varying levels of previous experience and the interactive nature of the lessons.

The second question explores the role of voluntary self-training during the summer vacation. Previous research has shown that learners can improve their oral proficiency through self-directed training (Foote & McDonough, 2017; Kuniyama et al., 2022; Nakanishi et al., 2022). It is hypothesized that higher participation in the self-training tasks, such as shadowing and other related activities, will lead to greater improvements in participants' oral proficiency. The study explored whether consistent and intensive engagement with self-training can effectively enhance learners' ability to produce fluent, accurate, and contextually appropriate spoken English.

This study examines the changes in learners' oral proficiency using four sub-scores of the Versant Speaking Test: Fluency, Pronunciation, Sentence Mastery, and Vocabulary, taken in April, July, and September.

Method

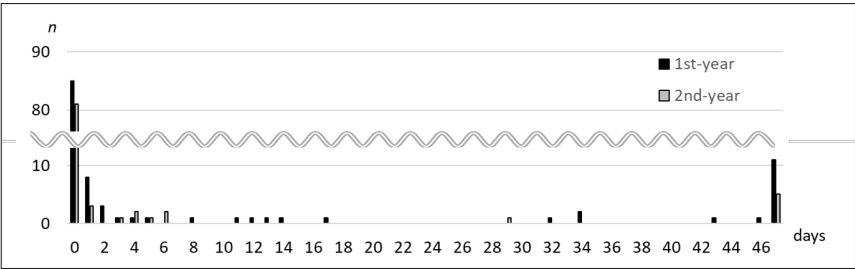
Participants

The participants were students enrolled in mandatory English conversation courses at a Japanese university during the spring semester of the 2023 academic year. These courses followed separate syllabi for first- and second-year students. Classes were randomly assigned and consisted of 20 to 25 students, meeting twice a week for 15 weeks. At the start of the semester, students' oral English proficiency ranged from CEFR A1 to A2, as measured by the Versant English Speaking Test (Pearson, 2022). Due to the university's requirement for third-year students to study abroad, most participants were relatively motivated to improve their English proficiency.

The first part of this study examined changes in oral English proficiency over the course of the semester among first-year ($n = 121$) and second-year ($n = 96$) students. All students took the Versant English Speaking Test in April (beginning of the semester), July (end of the semester/beginning of summer vacation), and September (end of summer vacation).

During the final week of the spring semester, students were invited to participate in an optional self-training program over the summer vacation. Participation was voluntary and unrelated to students' grades in either the spring or upcoming fall semester. Figure 1 illustrates the number of students (vertical axis) by the number of days they engaged in self-training out of the total 47-day period (horizontal axis). There were 166 students who had 0 days of participation, meaning they did not engage in self-training even once. A total of 51 students completed at least one training session. Among these, the data showed a clear divide between lower participation (1–17 days, $n = 29$) and higher participation (29–47 days, $n = 22$). In the second part of this study, changes in oral proficiency were analyzed across these three groups: "Never" ($n = 166$), "Lower" ($n = 29$), and "Higher" ($n = 22$) participation groups.

Figure 1
Number of First- and Second-Year Students by Total Days of Participation



Procedures

In-Class and Homework Activities During the Spring Semester

The English conversation courses followed a flipped learning model (Lee & Wallace, 2018). Drawing on the instructional cycle proposed by Nakanishi et al. (2019), students completed preparatory homework assignments, such as listening comprehension, shadowing-related tasks, oral presentation script writing, and recorded presentation rehearsals, prior to attending class. These tasks were designed to support active use of English during in-class activities. Classroom sessions focused on oral output, exposure to authentic input, and opportunities for self-monitoring and peer interaction. After in-class activities, students uploaded their presentation videos to YouTube, provided peer feedback to their classmates' videos, and reflected on their own work. One instructional cycle spanned approximately two to three weeks. All classroom instruction and discussions were conducted in English, while homework instructions, which were intended for independent study, were provided in Japanese to ensure clarity and minimize confusion.

These assignments and activities aligned closely with the classroom discussion topics. Table 1 presents the discussion topics covered in the spring semester. First-year students worked with a textbook that focused on familiar, everyday topics related to college life (Nakanishi et al., 2023a), and were encouraged to use simple English expressions to become comfortable with oral communication. For most students, this was their first experience in an interactive, communication-focused English class, as many Japanese secondary school classes tend to emphasize reading and grammar in large-group settings (Teeter, 2017). Second-year students used a textbook centered on social issues (Nakanishi et al., 2023b). These

students had completed the English conversation course during the previous year and were expected to develop and express their opinions more logically. Both first- and second-year classes met twice weekly. Each unit topic was practiced over two to three weeks, with Friday classes emphasizing interactive discussion and Thursday classes focusing on presentation skills (e.g., pair work, group work, rehearsals).

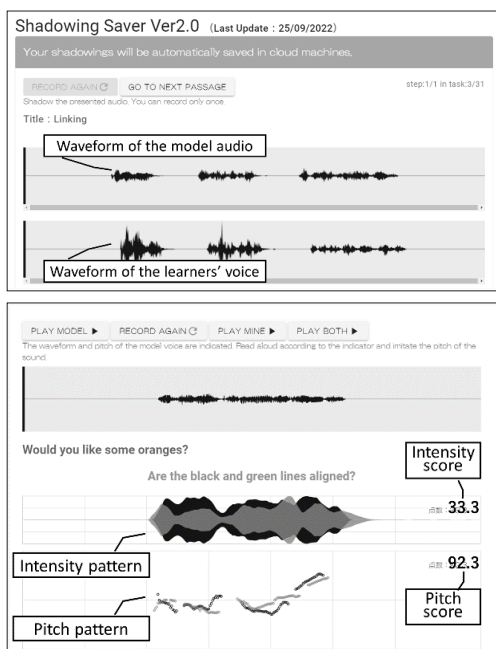
Table 1
Discussion Topics for the Spring Semester

Unit	1st year	2nd year
1	College life	Academic research
2	Understanding copyright	Social issues in Japan
3	Cyberbullying	Personal safety
4	Tourism	Gender
5	Foreign encounters	Religion
6	Entertainment	Business

Note. Excerpted from Nakanishi et al. (2023a, 2023b).

Intensive Self-Training During the Summer Vacation

During the 47-day summer vacation, an online self-training program was offered, with shadowing-based tasks as a central component following the procedure introduced in Minematsu et al. (2023). Figure 2 shows how shadowing and overlapping tasks were presented on the training platform. In the shadowing task, learners repeated what they heard through their headphones, and the waveform of their utterances was displayed beneath the waveform of the model audio. In the overlapping task, learners spoke in unison with the model voice while monitoring the intensity and pitch patterns displayed on the screen. Learners could repeat overlapping tasks to improve their score, which was based on the degree of match in pitch and intensity patterns.

Figure 2*Interface of Shadowing and Overlapping Tasks on the Training Platform*

Note. The top screen shows the shadowing task, where students repeat aloud what they hear through headphones. The top waveform displays the model voice, and the bottom waveform shows the learner's shadowed voice. The bottom screen shows the overlapping task, where learners speak simultaneously with the model audio. The interface provides real-time feedback on pitch and intensity alignment, displayed as correlation scores. Learners may repeat the task to improve their performance.

As summarized in Table 2, the training program was divided into three phases: Phase I (first 10 days) focused on connected speech, Phase II (30 days) on daily expressions, and Phase III (last 7 days) on prosodic features. All audio materials and transcripts were sourced from Nakanishi (2023). Participants were instructed to record their voice for each task and submit daily activity reports via a dedicated website. They were expected to spend at least 30 minutes on the seven daily tasks. Upon submission of their reports, participants provided consent for their recordings to be used for educational and research purposes under conditions of anonymity.

Table 2
Tasks Assigned for Intensive Self-Training During Summer Vacation 2023

Phase	Focus
I. Aug. 4– Aug. 13	Three sentences per day, focusing on features of connected speech (e.g., linking, elision, assimilation, and reduction).
II. Aug. 14– Sept. 12	Five sentences per day, focusing on expressions frequently used in daily communication.
III. Sept. 13– Sept. 19	One-minute excerpts from celebrity speeches, focusing on rhythm, intonation, speech rate, and emotion.
Task	Description
1) Shadowing (twice, normal speed)	Without script display, learners listen to the model audio and simultaneously repeat it aloud. The same utterance is shadowed and recorded twice. (Figure 2, top)
2) T/F Questions	Learners read sentences displayed on the screen and determine whether they match the content of the audio in Task 1 to check comprehension.
3) Synchro-Reading (normal speed)	With the script visible, learners read aloud in sync with the model audio, guided by a waveform indicator. Listening accuracy is calculated by comparing recordings from Tasks 1 and 3.
4) Synchro-Reading (slow audio)	Same as Task 3, but using slowed-down audio (0.5x and 0.8x speeds). Learners focus more closely on sound changes in connected speech.
5) Synchro-Reading (normal speed)	Repeating Task 3 at normal speed. Learners may re-record as many times as desired to improve pronunciation.
6) Overlapping (normal speed)	The script and real-time visual feedback (intensity and pitch patterns) are provided. Learners speak in unison with the model voice, aiming to exceed 60% in both pitch and intensity correlation scores. (Figure 2, bottom)
7) Reading Aloud	Without audio support, learners read the script aloud while following a visual pacing indicator. This task allows learners to check the final quality of their pronunciation and rhythm.

Versant English Speaking Test

The Versant English Speaking Test evaluates learners' spoken English proficiency using five task types: Reading, Repeating, Short Answer Questions, Sentence Building, and Story Retelling (Pearson, 2022). It provides an Overall score and four diagnostic sub-scores:

Fluency: Assesses rhythm, phrasing, and timing in sentence construction, reading, and repetition.

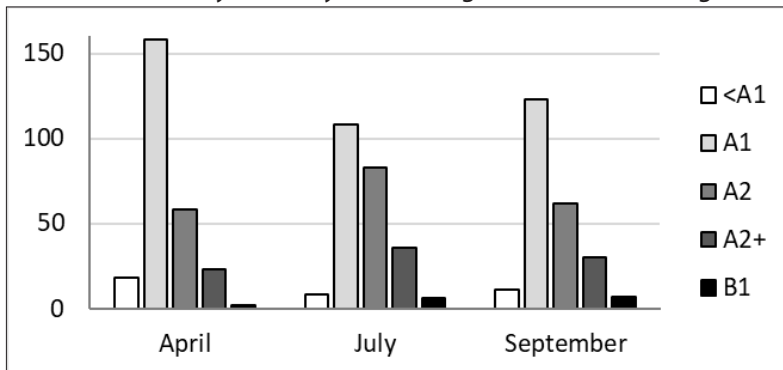
Pronunciation: Measures native-like production of consonants, vowels, and stress within sentence contexts.

Sentence Mastery: Reflects the ability to produce syntactically accurate and meaningful sentences using appropriate word and clause structures.

Vocabulary: Indicates familiarity with the form, meaning, and contextual use of everyday words in connected speech.

Overall and sub-scores range from 10 to 90 on Pearson's Global Scale of English (GSE), which is aligned with the CEFR (Council of Europe, 2001). The test's reliability and construct validity are supported by prior research (De Jong et al., 2016; Mayor et al., 2016). Downey et al. (2008) applied Bachman and Palmer's (1996) framework to evaluate test usefulness and found the Versant Test to have high authenticity and validity. Isbell and Kremmel (2020) reported moderate to high correlations between Versant scores and other standardized tests, such as IELTS and TOEFL iBT. Okamura (2013) found that Pronunciation and Vocabulary scores in particular affected intelligibility in Japanese learners' speech.

Students took the test three times: in April (start of semester), July (end of semester/start of vacation), and September (end of vacation). The April–July change was used to examine the effects of classroom learning, while the July–September change reflected developments during the summer vacation. Students who repeated the same first- or second-year course from the previous year were excluded. Only students who completed all three test sessions were included in the analysis. Figure 3 presents the distribution of students by CEFR levels at each testing period.

Figure 3*Student Proficiency Levels by CEFR Categories at Each Testing Period*

Note. The correspondence between the CEFR level and the Versant Overall GSE score is <A1 = 21 or below, A1 = 22–29, A2 = 30–35, A2+ = 36–42, and B1 = 43–50.

Results

Changes in EFL Learners' Oral Proficiency Over Six Months

To address RQ1, changes in learners' oral proficiency were examined over a six-month period in 2023. Versant Speaking Test scores were collected at three points: the beginning of the spring semester (April), the end of the semester (July), and the end of the summer vacation (September). First-year ($n = 121$) and second-year ($n = 96$) students were analyzed separately due to differences in their syllabi. Two-way ANOVAs were conducted on each Versant sub-score, with Year Group (1st, 2nd) as a between-subjects factor and Month (Apr, Jul, Sep) as a within-subjects factor, using Langtest (Mizumoto, 2015).

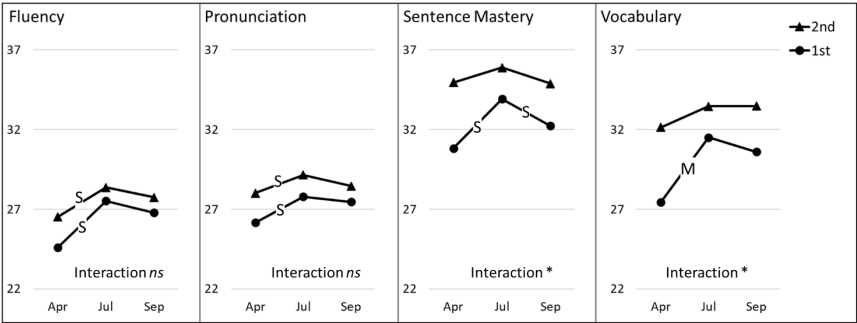
Figure 4 illustrates sub-score changes by Month and Year Group. Most mean scores were within the CEFR A1 (GSE 22–29) to A2 (GSE 30–35) range. Two-way ANOVAs revealed no statistically significant interaction between Month and Year Group for Fluency [$F(2, 430) = 1.43, p = .241, \eta^2 p = .007$] or Pronunciation [$F(2, 430) = 1.05, p = .351, \eta^2 p = .005$]. However, interactions for Sentence Mastery [$F(2, 430) = 4.41, p = .013^*, \eta^2 p = .020$] and Vocabulary [$F(2, 430) = 4.22, p = .015^*, \eta^2 p = .019$] were significant, prompting further analyses by Month and Year Group.

Regarding group differences, the main effect of Year Group was not significant for Fluency [$F(1, 215) = 1.83, p = .178, \eta^2 p = .008$], and marginal for Pronunciation [$F(1, 215) = 3.69, p = .056^+, \eta^2 p = .017$]. Sentence Mastery scores were significantly higher for second-year students in both

April [$F(1, 215) = 24.52, p < .001^{***}, \eta^2p = .102$] and September [$F(1, 215) = 7.38, p = .007^{**}, \eta^2p = .033$]. Similarly, their Vocabulary scores were significantly higher in April [$F(1, 215) = 29.53, p < .001^{***}, \eta^2p = .121$] and September [$F(1, 215) = 6.87, p = .009^{**}, \eta^2p = .031$], indicating that second-year students outperformed first-year students in these sub-scores.

Main effects of Month and simple main effects of Month by Year Group were further examined. Holm’s sequentially rejective Bonferroni procedure was applied for multiple comparisons, and effect sizes (Cohen’s d) were computed (see Table 3).

Figure 4
Versant Speaking Test Mean Scores by Month and Year Group



Note. Data were collected in April (Apr), July (Jul), and September (Sep). Triangles represent second-year students ($n = 96$), and circles represent first-year students ($n = 121$). S = small ($d > .20$); M = medium ($d > .50$).

Table 3
Effects of Month by Year Group (Cohen’s d).

	Fluency		Pronunciation		Sentence Mastery		Vocabulary	
Year	Apr –Jul	Jul –Sep	Apr –Jul	Jul –Sep	Apr –Jul	Jul –Sep	Apr –Jul	Jul –Sep
First	.39 ↗	.09	.28 ↗	.05	.49 ↗	.24 ↘	.58 ↗	.12
Second	.26 ↗	.08	.20 ↗	.11	.15	.15	.19	.00

Note. Apr = April; Jul = July; Sep = September.

The symbol “↗” indicates a positive effect size, while “↘” indicates a negative effect size during the given period. Note that arrows refer to effect size direction, not necessarily statistical significance.

Fluency

Fluency scores improved during the semester and slightly declined during the summer vacation. A significant main effect of Month was found for Fluency [$F(2, 430) = 24.72, p < .001^{***}, \eta^2p = .103$]. Scores improved from April ($M = 25.6, SD = 6.84$) to July ($M = 27.9, SD = 7.91$) [adj. $p < .001^{***}$], with small effect sizes both for first-year ($d = .39$) and second-year ($d = .26$) students. From July to September, scores slightly declined ($M = 27.3, SD = 7.36$) [adj. $p = .049^*$], though the associated effect sizes were minimal ($d = .09$ for first-years, $d = .08$ for second-years). These results suggest that fluency gains were primarily achieved during the semester, but did not continue through the summer break.

Pronunciation

Pronunciation scores increased during the semester and then remained stable. A significant main effect of Month was found for Pronunciation [$F(2, 430) = 11.17, p < .001^{***}, \eta^2p = .049$]. Scores improved from April ($M = 27.1, SD = 5.60$) to July ($M = 28.5, SD = 6.26$) [adj. $p < .001^{***}$], with small effect sizes ($d = .28$ for first-years; $d = .20$ for second-years). A marginal decline was observed in September ($M = 28.0, SD = 5.94$) [adj. $p = .074^+$] with minimal effect sizes ($d = .05$ and $.11$, respectively). Overall, pronunciation showed modest gains during coursework and remained relatively stable during the break, with effect sizes mostly falling below the threshold for small effects.

Sentence Mastery

The development of sentence mastery was evident primarily in first-year students. Simple main effects of Month were significant for first-year students [$F(2, 240) = 17.10, p < .001^{***}, \eta^2p = .125$] and marginal for second-years [$F(1.91, 181) = 2.46, p = .091^+, \eta^2p = .025$]. First-year scores increased significantly from April ($M = 30.8, SD = 6.22$) to July ($M = 33.9, SD = 6.44$) [adj. $p < .001^{***}, d = .49$, small effect], then declined in September ($M = 32.2, SD = 7.44$) [adj. $p = .002^{**}, d = .24$, small negative effect]. On the other hand, second-year students showed no significant change with negligible effect sizes: April ($M = 35.0, SD = 5.96$), July ($M = 35.9, SD = 6.22$) [adj. $p = .120, d = .15$], and September ($M = 34.9, SD = 6.76$) [adj. $p = .135, d = .15$]. These findings suggest that grammatical development was observed in first-years during coursework but partially regressed over the break, while second-years showed a plateau pattern.

Vocabulary

Vocabulary improvement during the semester was notable among first-year students. A significant main effect of Month was found for Vocabulary in the first-year group [$F(2, 240) = 24.51, p < .001^{***}, \eta^2p = .170$], but not for second-years [$F(2, 190) = 2.02, p = .136, \eta^2p = .021$]. First-year scores increased significantly with a medium effect size from April ($M = 27.5, SD = 6.29$) to July ($M = 31.5, SD = 7.38$) [adj. $p < .001^{***}, d = .58$], then remained statistically unchanged in September ($M = 30.6, SD = 7.96$) [adj. $p = .108, d = .12$]. Second-year scores remained consistent across all time points: April ($M = 32.1, SD = 6.35$), July ($M = 33.5, SD = 8.41$), and September ($M = 33.5, SD = 8.06$). This pattern suggests that vocabulary gains occurred primarily among first-year students during the semester and were retained during the break, while second-year students' performance remained stable.

Effects of Intensive Self-Training During Summer Vacation

To address RQ2, the effects of intensive self-training during summer vacation on learners' oral proficiency were analyzed. Students were grouped according to their training participation rate, regardless of year group, as the same self-training tasks were assigned across both cohorts. Participants who engaged in self-training for 29 days or more ($n = 22$) were classified as the "Higher" group, those who participated for 17 days or fewer ($n = 29$) as the "Lower" group, and those who did not engage in training at all ($n = 166$) as the "Never" group. Given the unequal group sizes and unbalanced variances, non-parametric tests were employed to evaluate changes in Versant sub-scores over time and across groups.

To determine whether students with higher initial proficiency were more likely to participate in summer self-training, group differences in each test month were examined using Kruskal-Wallis tests. As shown in Table 4, no significant differences were observed among the three groups in April and July, suggesting similar proficiency levels at the beginning and end of the spring semester. However, by September, significant differences emerged in Fluency, Pronunciation, and Sentence Mastery sub-scores. Subsequent post hoc pairwise comparisons using the Mann-Whitney U test with Holm-Bonferroni correction indicated that in September, the Higher group outperformed the Lower and Never groups in Fluency, Pronunciation, and Sentence Mastery, with small to medium effect sizes. Comparing the Lower and Never groups in September, significant differences were observed only in Pronunciation.

Table 4
Kruskal-Wallis Tests on Group Differences in Versant Sub-scores by Month (df = 2)

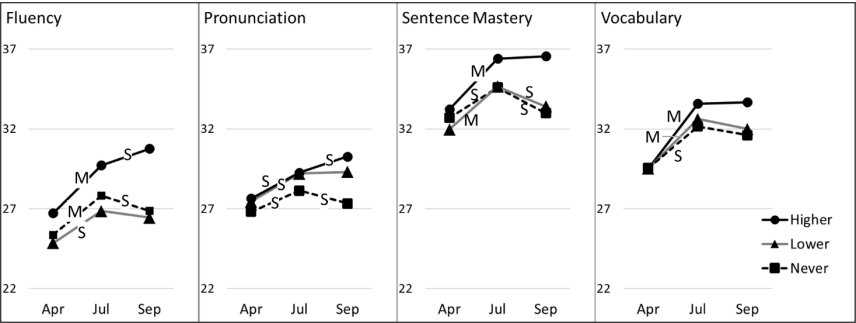
	Fluency			Pronunciation			Sentence Mastery			Vocabulary		
	Apr	Jul	Sep	Apr	Jul	Sep	Apr	Jul	Sep	Apr	Jul	Sep
χ^2	1.54	2.26	6.85	2.34	1.11	11.72	0.64	1.27	5.19	0.10	1.38	0.97
p	.463	.323	.033*	.310	.573	.003**	.725	.530	.075+	.953	.501	.617
r	.050	.067	.145	.069	.038	.202	.024	.043	.121	.004	.046	.034

Note. Apr = April; Jul = July; Sep = September.

+ $p < .10$; * $p < .05$; ** $p < .01$.

Figure 5 illustrates the trends in mean Versant Speaking Test scores by Month (Apr, Jul, Sep) and Participant Group (Higher, Lower, Never). Friedman rank sum tests were conducted to examine longitudinal changes in scores within each group. Statistically significant differences were found in all three groups across months. For post hoc comparisons, Wilcoxon signed-rank tests with Holm-Bonferroni correction were performed between months (April–July, July–September) and effect size r was calculated, as presented in Table 5.

Figure 5
Versant Speaking Test Mean Scores by Month and Participant Group



Note. Scores were measured in April (Apr), July (Jul), and September (Sep). The solid line with circles represents the Higher group ($n = 22$), the gray line with triangles the Lower group ($n = 29$), and the dotted line with squares the Never group ($n = 166$).
S = small ($r > .10$); M = medium ($r > .30$).

Table 5

Wilcoxon Signed-Rank Tests for Within-Group Changes Over Time (Effect Size r)

Year	Fluency		Pronunciation		Sentence Mastery		Vocabulary	
	Apr -Jul	Jul -Sep	Apr -Jul	Jul -Sep	Apr -Jul	Jul -Sep	Apr -Jul	Jul -Sep
Higher	.435 ↗	.175 ↗	.274 ↗	.284 ↗	.465 ↗	.068	.359 ↗	.037
Lower	.287 ↗	.096	.283 ↗	.017	.303 ↗	.162 ↘	.394 ↗	.065
Never	.302 ↗	.114 ↘	.211 ↗	.148 ↘	.245 ↗	.226 ↘	.225 ↗	.050

Note. Apr = April; Jul = July; Sep = September.

The symbol “↗” indicates a positive effect size, while “↘” indicates a negative effect size during the given period. Note that arrows refer to effect size direction, not necessarily statistical significance.

Fluency

Fluency scores improved during the spring semester across all groups, but diverged during the summer depending on the level of self-training. The Higher group showed an increase from April ($M = 26.7$, $SD = 5.95$) to July ($M = 29.7$, $SD = 7.62$), with a continued rise in September ($M = 30.8$, $SD = 7.03$). The April–July improvement was statistically significant ($z = 2.884$, $\text{adj.}p = .008^*$, $r = .435$), indicating a medium effect, while the July–September gain was not significant ($z = 1.163$, $\text{adj.}p = .245$, $r = .175$), but reflecting a small effect. The Lower group also improved from April ($M = 24.9$, $SD = 6.67$) to July ($M = 26.9$, $SD = 7.47$), followed by a slight drop in September ($M = 26.5$, $SD = 6.76$). The semester gain was marginally significant ($z = 2.185$, $\text{adj.}p = .087^+$, $r = .287$), suggesting a small effect, while the summer decline was negligible ($z = 0.730$, $\text{adj.}p = .465$, $r = .096$). The Never group improved from April ($M = 25.4$, $SD = 6.99$) to July ($M = 27.8$, $SD = 8.03$), but scores decreased in September ($M = 26.9$, $SD = 7.41$). The April–July gain was statistically significant ($z = 5.502$, $\text{adj.}p < .001^{***}$, $r = .302$), indicating a medium effect, and the summer drop was also significant ($z = 2.081$, $\text{adj.}p = .037^+$, $r = .114$), representing a small negative effect. Overall, fluency gains were best sustained by the Higher group, who engaged in frequent self-training.

Pronunciation

All groups showed gains in pronunciation during the semester, but their trajectories varied over the summer. The Higher group moved from April ($M = 27.6$, $SD = 5.21$) to July ($M = 29.3$, $SD = 6.67$), and then to September ($M = 30.3$, $SD = 4.34$). The gains in mean scores were not statistically significant either during the semester ($z = 1.817$, $\text{adj.}p = .119$, $r = .274$) or over the summer ($z = 1.885$, $\text{adj.}p = .119$, $r = .284$), but small effect sizes were indicated for both periods. The Lower group improved from April ($M = 27.5$, $SD = 5.69$) to July ($M = 29.2$, $SD = 6.18$), and held steady in September ($M = 29.3$, $SD = 6.07$). The semester gain was marginally significant ($z = 2.155$, $\text{adj.}p = .092^+$, $r = .283$, small effect), and no significant change was observed in the summer ($z = 0.133$, $\text{adj.}p = .894$, $r = .017$). The Never group improved from April ($M = 26.8$, $SD = 5.66$) to July ($M = 28.1$, $SD = 6.24$), but declined by September ($M = 27.3$, $SD = 6.01$). The spring semester gain was significant ($z = 3.847$, $\text{adj.}p < .001^{***}$, $r = .211$, small effect), and the summer drop was also significant ($z = 2.702$, $\text{adj.}p = .014^+$, $r = .148$, small negative effect). The results show that the Higher group was more successful than the other groups in terms of pronunciation scores throughout the semester and summer vacation.

Sentence Mastery

Sentence mastery improved during the semester for all groups, but only the Higher group maintained their mean score through the summer. The Higher group progressed from April ($M = 33.2$, $SD = 5.80$) to July ($M = 36.4$, $SD = 5.66$), with performance remaining stable in September ($M = 36.6$, $SD = 5.90$). The April–July improvement was significant ($z = 3.084$, $\text{adj.}p = .006^{**}$, $r = .465$), showing a medium effect, and the July–September change was negligible ($z = 0.449$, $\text{adj.}p = .653$, $r = .068$). The Lower group increased from April ($M = 32.0$, $SD = 6.13$) to July ($M = 34.7$, $SD = 5.78$), but then declined to September ($M = 33.4$, $SD = 5.75$). The semester gain was marginally significant with a medium effect ($z = 2.304$, $\text{adj.}p = .064^+$, $r = .303$). The summer decrease was not significant, but a small effect size was shown ($z = 1.233$, $\text{adj.}p = .218$, $r = .162$). The Never group improved from April ($M = 32.7$, $SD = 6.58$) to July ($M = 34.6$, $SD = 6.60$), followed by a drop in September ($M = 33.0$, $SD = 7.57$). The spring gain was statistically significant ($z = 4.465$, $\text{adj.}p < .001^{***}$, $r = .245$, small effect), and the summer decline was also significant ($z = 4.111$, $\text{adj.}p < .001^{***}$, $r = .226$), again a small effect. In sum, only the Higher group was able to sustain sentence mastery gains over the break.

Vocabulary

Vocabulary improved significantly for all groups during the semester, with minimal change over the summer. The Higher group rose from April ($M = 29.5$, $SD = 5.47$) to July ($M = 33.6$, $SD = 6.74$), and remained stable in September ($M = 33.7$, $SD = 9.23$). The semester gain was marginally significant ($z = 2.380$, $\text{adj.}p = .052^+$, $r = .359$, medium effect), and no statistical change was seen in the summer ($z = 0.244$, $\text{adj.}p = .808$, $r = .037$). The Lower group increased from April ($M = 29.5$, $SD = 7.06$) to July ($M = 32.6$, $SD = 6.39$), then declined slightly in September ($M = 32.0$, $SD = 8.68$). The April–July gain was statistically significant ($z = 3.002$, $\text{adj.}p = .008^*$, $r = .394$, medium effect), and the summer change was not significant ($z = 0.493$, $\text{adj.}p = .622$, $r = .065$). The Never group moved from April ($M = 29.5$, $SD = 6.85$) to July ($M = 32.2$, $SD = 8.28$), then to September ($M = 31.6$, $SD = 7.87$). The semester gain was significant ($z = 4.099$, $\text{adj.}p < .001^{***}$, $r = .225$), showing a small effect, while the summer decline was not significant ($z = 0.916$, $\text{adj.}p = .360$, $r = .050$). It was suggested that vocabulary gains occurred during the semester and were retained over the summer in all groups.

Discussion

The Role of Classroom Interaction in Oral Proficiency Development in EFL Contexts (RQ1)

This study first examined how learners' oral proficiency changed across the spring semester and the subsequent summer vacation, comparing first- and second-year students. The results revealed the pedagogical value of structured, interactive classroom activities in fostering core aspects of spoken English performance in EFL contexts such as Japan. In particular, improvements were observed in dimensions closely tied to real-time speech production (fluency and pronunciation) while development in grammar and vocabulary showed a more differentiated pattern by year group. These findings provide insight into how different components of oral proficiency may respond differently to classroom-based instruction and to subsequent absence of input.

For both first- and second-year students, fluency and pronunciation scores improved during the semester. These gains, supported by small effect sizes, point to the impact of interaction-focused lessons in facilitating smoother and more intelligible speech. However, during the summer break, learners' gains in both domains stagnated. This pattern reinforces the notion that

regular opportunities for output and feedback, as provided by classroom interaction, are essential for developing and sustaining prosodic and articulatory fluency (Lightbown & Spada, 2013; Zhang, 2009). The absence of such interaction may hinder learners' improved performance in these areas, due to reduced procedural use of language skills (Hakuta et al., 2000).

Sentence mastery and vocabulary, by contrast, showed gains only among first-year students during the semester, with no notable change observed among second-year learners. This discrepancy likely reflects the difference in baseline oral proficiency, as second-year students had already completed one year of similar coursework. The improvements observed in first-year students may reflect the effectiveness of instructional input in consolidating morphosyntactic and lexical knowledge during earlier stages of acquisition (Muñoz, 2014). However, these gains were not fully retained during the vacation period. Sentence mastery declined slightly among first-years and remained flat among second-years, while vocabulary scores remained statistically stable across both groups. This suggests that beyond the initial stages of acquisition, both sentence formation and vocabulary breadth may require targeted instruction and recycling to facilitate sustained improvement (Trofimovich et al., 2009; Nation, 2007).

In summary, the results from RQ1 support the effectiveness of interactive classroom instruction in promoting short-term improvements in fluency and pronunciation. However, morphosyntactic and lexical skills may require different pedagogical inputs and may be more vulnerable to regression in the absence of continued engagement. The findings also highlight the difficulty of maintaining oral proficiency during instructional breaks in EFL environments where learners have limited access to authentic communication opportunities (Teeter, 2017).

The Impact of Intensive Self-Training During Summer Vacation (RQ2)

The second part of the study investigated the extent to which intensive self-training during the summer vacation could help learners maintain or enhance their oral proficiency. While the three groups (Higher; Lower; and Never) showed comparable progress during the spring semester, their trajectories diverged over the summer break. This section focuses on these differences during the vacation period, when classroom instruction was absent and learners' engagement with self-training varied.

With regard to fluency, learners in the Higher participation group (29–47 days of training) demonstrated continued improvement from July to

September. Although not statistically significant, the notable effect size was observed, suggesting that regular practice may have supported incremental gains. The Never group, by contrast, experienced statistically detectable declines, albeit with small effect sizes, while the Lower group (1–17 days of training) remained relatively stable. These findings suggest that the frequency and consistency of self-training influenced learners' ability to sustain fluency, supporting the idea that proceduralized language skills require ongoing rehearsal (Foote & McDonough, 2017; Kadota, 2019).

Pronunciation followed a similar pattern. The Higher group showed small gains over the summer, though not statistically significant. In contrast, the Never group experienced statistically significant decline, and the Lower group exhibited minimal change. These patterns imply that consistent self-practice, particularly through shadowing, may help reinforce articulatory habits and prosodic features that are otherwise prone to attrition during breaks in instruction (Cooper, 2003). While lower levels of training did not lead to gains, they may have served to buffer against substantial decline.

Sentence mastery showed clear differentiation among the groups. Only the Higher group was able to maintain their performance over the summer. The non-significant decline over the summer observed in the Higher group suggests that even though shadowing is not explicitly focused on grammar, repeated exposure to and production of grammatical structures may contribute to the retention of sentence-level accuracy. Conversely, both the Lower and Never groups exhibited declines with small effect sizes. Insufficient practice, as seen in the Lower and Never groups, appears to leave learners vulnerable to regression in this domain.

In the case of vocabulary, all groups remained relatively stable during the summer break, with no statistically significant changes and negligible effect sizes. This stability suggests that vocabulary knowledge, particularly of high-frequency words used in everyday contexts, may be more resistant to short-term loss compared to skills like fluency or pronunciation, which are more procedural in nature. Alternatively, the absence of further gains indicates that vocabulary development may require explicit instruction and deliberate recycling (Nation, 2007), conditions not fully present in shadowing-based training.

Taken together, these findings indicate that intensive self-training, when conducted consistently and with sufficient frequency, may help preserve and even enhance certain dimensions of oral proficiency during breaks from formal instruction. While limited training can offer some protective

effect, the observed differences across groups indicate the importance of sustained engagement. In EFL contexts where learners lack naturalistic exposure, structured self-training activities, especially those involving real-time input-output processing (Kadota, 2019), can serve as a viable supplement to classroom-based learning. Encouraging autonomy through accessible and rhythmically supported tools such as shadowing may thus be a valuable strategy for long-term oral development.

Conclusion and Limitations

This study examined the temporal dynamics of EFL learners' oral proficiency by analyzing changes before and after the spring semester, as well as after the summer vacation, focusing on the roles of coursework and self-directed practice using shadowing-based tasks. Using Versant Speaking Test sub-scores (Fluency, Pronunciation, Sentence Mastery, and Vocabulary), the study tracked proficiency changes among 217 university students in two different academic years (first-year and second-year cohorts). The findings offer two key insights into how language skills develop and fluctuate depending on learners' engagement in classroom activities and autonomous practice.

First, the results confirmed that interactive, communication-focused classroom instruction contributes to measurable improvements in learners' oral fluency and pronunciation over a single semester. These short-term gains were notable in first-year students, likely due to their limited prior exposure to spoken English and the novelty of the flipped learning format. However, without sustained engagement during the summer break, learners' proficiency tended to stagnate or regress, particularly in fluency, pronunciation, and syntactic accuracy, suggesting the importance of consistent language use in maintaining oral skills.

Second, the study found that self-directed practice involving shadowing and related oral tasks during academic breaks can serve as a partial buffer against such declines. Learners who practiced more intensively tended to maintain or even slightly improve their performance in fluency, pronunciation, and sentence mastery, while those who practiced less frequently experienced minimal gains or observable regression. These outcomes emphasize the potential of structured, repetitive oral training as a means of sustaining procedural language skills outside the classroom.

From a pedagogical perspective, these findings support the integration of interactive classroom activities with accessible tools for autonomous practice. Instructors and curriculum designers should consider embedding

rhythmically supported, output-based tasks like shadowing into both in-class and out-of-class learning opportunities. Moreover, institutions should provide explicit support for learner autonomy, particularly during long academic breaks when instructional input is limited.

At the same time, some limitations should be noted. First, this study did not control for participants' use of English outside the classroom, such as part-time jobs involving English or interaction with international students. Such factors may have influenced individual outcomes and should be more systematically accounted for in future research. Second, the grouping criteria for summer self-training were based on participation frequency (never; 1–17 days; 29–47 days), which, while practically motivated, involved a degree of arbitrariness. Treating participants who trained once and those who trained 17 times as a single group may obscure meaningful within-group variation. Third, the study compared learners across two academic years but did not follow individuals longitudinally, making it difficult to draw conclusions about long-term developmental trajectories. Future studies should investigate the qualitative nature of training engagement and explore more nuanced ways to capture its frequency, consistency, and intensity.

Despite these limitations, the findings contribute to our understanding of oral proficiency development in EFL contexts. The temporal dynamics of learners' oral proficiency are shaped not only by formal instruction but also by their engagement during instructional gaps. With appropriate design and institutional support, self-training approaches, especially those centered around structured oral repetition, can help EFL learners sustain and build their speaking skills beyond the classroom.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Numbers 22H00527, 23K17459, and 24K04175. The shadowing training interface used in this study, Shadowing Saver, was developed in collaboration with the research team led by Professor Nobuaki Minematsu at the Graduate School of Engineering, the University of Tokyo.

Noriko Nakanishi is a Professor in the Faculty of Global Communication at Kobe Gakuin University. Her research interests include English Phonetics, Sociolinguistics, and Education.

References

- Bachman, L., & Palmer, A. (1996). *Language testing in practice*. Oxford University Press.
- Bygate, M. (2001). Speaking. In R. Carter & D. Nunan (Eds.), *The Cambridge guide to teaching English to speakers of other languages* (pp. 14–20). Cambridge University Press. <https://doi.org/10.1017/CBO9780511667206.003>
- Cooper, H. (2003). *Summer learning loss: The problem and some solutions*. ERIC Clearinghouse on Elementary and Early Childhood Education.
- Council of Europe (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge University Press.
- De Jong, J. H. A. L., Mayor, M., & Hayes, C. (2016). Developing global scale of English learning objectives aligned to the common European framework. *Global Scale of English Research Series*.
- DeKeyser, R. M. (2007). *Practice in a second language: Perspectives from applied linguistics and cognitive psychology*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511667275>
- Downey, R., Farhady, H., Present-Thomas, R., Suzuki, M., & Van Moere, A. (2008). Evaluation of the usefulness of the Versant for English test: A response. *Language Assessment Quarterly*, 5(2), 160–167. <https://doi.org/10.1080/15434300801934744>
- Eyring, J. (2014). ESL and EFL learning environments. In M. Celce-Murcia, D. M. Brinton, & M. A. Snow (Eds.), *Teaching English as a second or foreign language* (4th ed., pp. 36–51). National Geographic Learning.
- Foote, J. A., & McDonough, K. (2017). Using shadowing with mobile technology to improve L2 pronunciation. *Journal of Second Language Pronunciation*, 3(1), 34–56. <https://doi.org/10.1075/jslp.3.1.02foo>
- Hakuta, K., Butler, Y. G., & Witt, D. (2000). How long does it take English learners to attain proficiency? *The University of California Linguistic Minority Research Institute Policy Report 2000*, 1.
- Hamada, Y. (2019a). Shadowing: What is it? How to use it. Where will it go? *RELC Journal*, 50(3), 386–393. <https://doi.org/10.1177/0033688218771380>
- Hamada, Y. (2019b). *Teaching shadowing for listening: Developing learners' bottom-up skills*. Routledge.
- Isbell, D. R., & Kremmel, B. (2020). Test review: Current options in at-home language proficiency tests for making high-stakes decisions. *Language Testing*, 37(4), 600–619. <https://doi.org/10.1177/0265532220943483>

- Kadota, S. (2019). *Shadowing as a practice in second language acquisition: Connecting inputs and outputs*. Routledge. <https://doi.org/10.4324/9781351049108>
- Kunihara, T., Zhu, C., Minematsu, N., & Nakanishi, N. (2022). Gradual improvements observed in learners' perception and production of L2 sounds through continuing shadowing practices on a daily basis. *INTERSPEECH Proceedings*, 1303–1307. <https://doi.org/10.21437/Interspeech.2022-101>
- Lambert, S. (1992). Shadowing. *Meta*, 37(2), 263–273. <https://doi.org/10.7202/003378ar>
- Lee, G., & Wallace, A. (2018). Flipped learning in the English as a foreign language classroom: Outcomes and perceptions. *TESOL Quarterly*, 52(1), 62–84. <https://doi.org/10.1002/tesq.372>
- Lightbown, P. M., & Spada, N. (2013). *How languages are learned* (4th ed.). Oxford University Press.
- Mayor, M., Seo, D., De Jong, J. H. A. L., & Buckland, S. (2016). *Technical report: Aligning CEFR-J descriptors to GSE*. <https://www.pearson.com/content/dam/one-dot-com/one-dot-com/pearson-languages/en-gb/pdfs/gse/gse-resources/gse-research-reports/technical-report-aligning-cefr-j-descriptors-to-gse.pdf>
- Minematsu, N., Nakanishi, N., Gao, Y., & Sun, H. (2023). A unified framework to improve learners' skills of perception and production based on speech shadowing and overlapping. *INTERSPEECH Proceedings*, 3667–3668.
- Ministry of Education, Culture, Sports, Science and Technology (MEXT). (1956). 大学設置基準 [Standards for the establishment of universities]. Retrieved from https://laws.e-gov.go.jp/law/331M50000080028#Mp-Ch_6
- Mizumoto, A. (2015). Langtest (Version 1.0) [Web application]. Retrieved from <http://langtest.jp>
- Muñoz, C. (2014). Contrasting effects of starting age and input on the oral performance of foreign language learners. *Applied Linguistics*, 35(4), 463–482. <https://doi.org/10.1093/applin/amu024>
- Nagle, C., Baese-Berk, M., & Trofimovich, P. (2025). High variability phonetic training: The role of talker variability and practice. *Language Learning*, 75(1), 142–170. <https://doi.org/10.1111/lang.12577>
- Nakanishi, N. (2023). *Eigo Shadowing 100pon Knock*. Cosmopier.
- Nakanishi, N., Minematsu, N., & Kunihara, T. (2022). Effects of English shadowing training using unlearned passages on listening and speaking skills. *Language Education & Technology*, 59, 77–105. https://doi.org/10.24539/let.59.0_77
- Nakanishi, N., Musty, N., Otake, S., Tam, S. Y., Ebihara, Y., & Fujimura, K. (2023a). *Global perspectives: Listening & speaking book 1*. Seibido.

- Nakanishi, N., Musty, N., Otake, S., Tam, S. Y., Ebihara, Y., & Fujimura, K. (2023b). *Global perspectives: Listening & speaking book 2*. Seibido.
- Nakanishi, N., Tam, S. Y., & Ebihara, Y. (2019). ICT教室の特性と音声認識ソフトを活かした英語リスニング・スピーキング活動の可視化 [Visualizing English listening and speaking activities using speech recognition software in ICT-enhanced classrooms]. *Japan Universities Association for Computer Education ICT Riyouniyoru Kyoiku Kaizen Kenkyu Happyokai*, 153–156.
- Nation, I. S. P. (2007). The four strands. *Innovation in Language Learning and Teaching*, 1(1), 2–13. <https://doi.org/10.2167/illt039.0>
- Okamura, A. (2013). Versant and advanced L2 speakers' ratings of Japanese learners' oral English. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 4th pronunciation in second language learning and teaching conference* (pp. 16–28). Iowa State University.
- Pearson. (2022). *Versant English speaking test: Test description and validation summary*. Pearson Education.
- Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 235–253). Newbury House.
- Teeter, R. (2017). Japanese university students' attrition in L2 listening proficiency: A case study. *JALT Journal*, 39(1), 25–43.
- Trofimovich, P., Lightbown, P. M., Halter, R. H., & Song, H. (2009). Comprehension-based practice: The development of L2 pronunciation in a listening and reading program. *Studies in Second Language Acquisition*, 31(4), 609–639. <https://doi.org/10.1017/S0272263109990040>
- Zhang, S. (2009). The role of input, interaction and output in the development of oral fluency. *English Language Teaching*, 2(4), 91–100. <http://doi.org/10.5539/elt.v2n4p91>