Articles

Exploring the Optimal Number of Repetitions for Shadowing: A Focus on Listening Comprehension, Memorization of Multiword Expressions, Bottom-Up Processing, and Repetition Speed

Ryotaro Hashizaki
Matsuyama University

Hirohisa Sekiyama
Kwansei Gakuin University

Yun Xing
Shanxi University

Kumi Wakita
Nagoya University

Shadowing enhances second-language learners' listening comprehension and the memorization of multiword expressions. However, it remains unclear whether the number of repetitions required varies based on different aspects of shadowing

https://doi.org/10.37546/JALTJJ47.1-1

effectiveness. This study involved 30 Japanese university students who participated in 11 shadowing training sessions, each lasting 90 minutes per week. Pre- and post-intervention tests measured their improvements in listening comprehension, multiword expression knowledge, bottom-up processing, and repetition speed. The number of shadowing repetitions completed both inside and outside the classroom was also recorded. The results showed that learners who completed more repetitions achieved higher post-test scores in listening comprehension and multiword expression knowledge. Statistically significant improvements were also observed in bottom-up processing and repetition speed between the pre- and post-tests, regardless of the number of repetitions. These findings indicate that the optimal number of repetitions depends on the specific learning targets when using shadowing. Additionally, the present study suggested the possibility that a stage of automating bottom-up processing may exist as a prerequisite for improving listening skills through shadowing.

シャドーイングは、第二言語学習者のリスニング理解および連語表現（multiword expressions, MWEs）の記憶を向上させる効果があるとされている。しかし、シャドーイングの効果の異なる側面で、必要な反復回数が異なるかどうか明らかでない。本研究では、日本人大学生30名を対象に、1回90分のシャドーイングトレーニングを週に1回、全11回実施した。トレーニング前後で、リスニング理解、MWEsの知識、ボトムアップ処理能力、および復唱速度の向上を測定し、教室内外で行われたシャドーイングの反復回数を記録した。結果、より多くの反復を行った学習者は、リスニング理解とMWEsの知識において高い事後テストスコアを獲得した。ボトムアップ処理能力と復唱速度に関しては、反復回数に関わらず、有意な向上がトレーニング後に確認された。これらの結果から、シャドーイングを用いた学習における最適な反復回数は、学習目標によって異なることが示唆された。さらに、シャドーイングによるリスニング能力向上の前提条件として、ボトムアップ処理能力の自動化の段階が存在する可能性が示された。

Keywords: bottom-up; EFL; listening; shadowing; vocalization

Enhancing English as a Foreign Language (EFL) learners' listening comprehension is important because it enables them to obtain the vast input of the target language (Vandergrift, 2007). Shadowing is a common method to enhance listening skills in an EFL environment. It is "an act or task of listening, in which the learner tracks the heard speech and repeats it as exactly as possible while listening attentively to the in-coming information" (Tamai, 2005, p. 34). Shadowing improves L2 listening skills by means of the enhancement of bottom-up processing. Bottom-up processing refers to the ability to "construct meaning by accretion, gradually combining increasingly larger units of meaning from the phoneme-level up to discourse-level features" (Vandergrift, 2004, p. 4). This is crucial because less-proficient learners often have deficiencies in their bottom-up processing skills (Field, 2003). Shadowing also fosters the memorization of multiword expressions (MWEs). MWEs denote expressions comprising multiple words and are critical as they facilitate fluent language usage including listening (Conklin & Schmitt, 2012). For instance, Tang (2013) suggests that learning English chunks improves L2 listening comprehension. While there are various terminologies used for expressions comprising multiple words (e.g., formulaic sequences, formulaic language, chunks, collocations, and idioms), this study employs MWEs because its focus is not on specific items but expressions comprising multiple words in general. Additionally, the improvement of repetition speed is the prerequisite for enhancing the bottom-up processing and the memorization of MWEs (Kadota, 2015, 2019). Although shadowing has been shown to be effective in improving L2 listening skills (Hamada, 2016a, 2016b; Tamai, 2005), there still remain questions on effective instructional methods of shadowing, such as how often shadowing should be repeated. In the remaining introduction, the effectiveness of shadowing, details of its mechanism, and the limitations of previous research will be discussed.

Effects of Shadowing and Its Theoretical Underpinnings

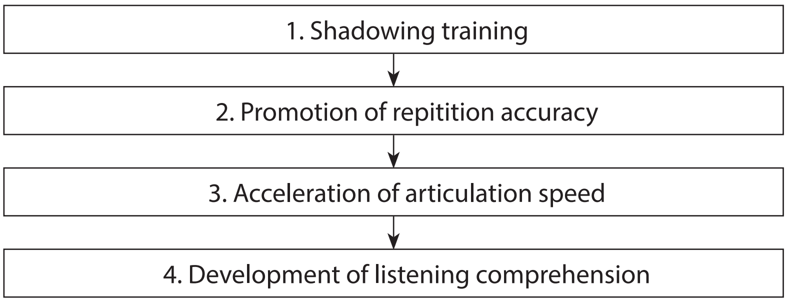
This study focuses on the effects of shadowing on the improvement of L2 listening comprehension (Hamada, 2016a; Tamai, 2005), and the memorization of MWEs (Hashizaki, 2021, 2024c; Miyake, 2009; Xing & Hashizaki, 2021, 2024). Additionally, L2 bottom-up processing (Hamada, 2016a), repetition speed (Tamai, 2005), and the theoretical underpinnings supporting the effect of shadowing, will be examined.

Regarding improvement in L2 listening comprehension, previous research (Hamada, 2016a; Tamai, 2005) has shown that shadowing is more effective for lower- than higher-proficiency learners (Hamada, 2016a; Tamai, 2005). Hamada (2016a) divided 43 Japanese undergraduate EFL students into two proficiency levels (low and intermediate) based on their initial test scores. After the training sessions, the participants took a listening test that consisted of two different question levels (basic and advanced). The results revealed that the lower-proficiency learners' basic listening test scores increased, while the intermediate-proficiency learners' scores did not. The advanced level listening test did not show a statistically significant improvement for the participants in both proficiency levels. Overall, these findings support the idea that engaging in shadowing exercises contributes to lower-proficient learners' development of L2 listening comprehension. Hamada (2016a) explained that "after training, low-proficiency learners approached the initial level of the intermediate group in terms of phoneme perception and listening comprehension" (p. 48).

Regarding its mechanism, shadowing makes learners pay attention to the sound of heard speech rather than its content (Hamada, 2016b; Kadota, 2019; O'ki, 2011), and this improves bottom-up processing. Based on Tamai's (2005) findings, Kadota (2019) argues that improving listening comprehension through shadowing involves a preliminary stage in which the ability to accurately repeat heard speech improves. For instance, Hamada (2016a) showed that shadowing training improves the ability to recognize words accurately measured with a dictation-cloze test. This is then followed by an improvement in articulation speed as observed in Tamai (2005) and Miyake (2009) (Figure 1).

Figure 1

Prerequisites for Improving Listening Comprehension Through Shadowing Based on Kadota (2019)



From the psycholinguistic perspective, van Paridon et al.'s (2019) model delineates two shadowing pathways: one involves processing meaning, and the other omits it. In the former pathway, learners undergo several linguistic steps before reproducing words, such as extracting phonetic features, segmenting, selecting phonological codes, and choosing lemma forms. Conversely, in the latter pathway, learners bypass lemma selection before reproducing what they heard. This implies that shadowing words does not guarantee an understanding of their meaning.

In addition to improving listening comprehension, shadowing is also effective for the memorization of MWEs. For example, Miyake (2009) showed that shadowing led to faster speech rates, which subsequently enabled learners to memorize MWEs. Hashizaki (2021, 2024c), and Xing and Hashizaki (2021, 2024) have also shown that shadowing leads to the memorization of MWEs included in passages.

Regarding the theoretical aspect of memorizing MWEs through shadowing, the underpinning theory is based on working memory (Kadota, 2015). Kadota (2015) reveals that shadowing facilitates the memorization of newly learned language items by improving the efficiency of rehearsal speed. The memory span of the phonological store, a component of working memory, is two seconds (Baddeley et al., 1975); therefore, the faster the repetition speed, the greater the amount of information that can be stored within two seconds and sent to one's long-term memory. Miyake (2009) suggests that MWEs that are vocalized within two seconds through repeated shadowing have significantly greater recall compared to those over two seconds, which implies that repeated shadowing accelerates articulation speed and fosters memorization. Similarly, Hashizaki (2021) uses passages as materials instead of phrases and reveals that MWEs that are repeated faster after shadowing training tend to be better memorized. Hashizaki (2024c) further suggests that the memorization of MWEs by shadowing has two routes: immediate and delayed. The immediate route is based on the findings of the production effect, which reveal that vocalization leads to better memory because it directs learners' attention to target items (Fawcett & Ozubko, 2016; Hashizaki, 2024a; Icht & Mama, 2022; MacLeod et al., 2010; Ozubko et al., 2012). The delayed route is in line with the explanation of Miyake (2009) and Kadota (2015); that is, shadowing leads to a faster repetition rate, which, in turn, induces the effective memorization of MWEs.

Effect of Repetitions on Improving Listening Comprehension by Shadowing

Suzuki (2023) states that automatization is driven by repetition, which is an integral part of practice. This is also the case for shadowing. Previous studies (Hamada, 2016a; O'ki, 2014; Shiki et al., 2010; Tamai, 2005) have stated that four to five repetitions are needed for effective shadowing for one material. According to O'ki (2014) and Shiki et al. (2010), four to five repetitions lead to a plateau in the shadowing reproduction rate (the ratio of correctly shadowed words or syllables). Accordingly, Hamada (2016a) uses this procedure to show that all learners improve their bottom-up processing, as measured by the dictation-cloze test. Additionally, Tamai (2005) found that repetition speed improves with shadowing with the volume of four to five repetitions for one material in one class.

Although the previous studies have shown the impact of four to five repetitions on the shadowing reproduction rate (O'ki, 2014; Shiki et al.,

2010) for one material, thus indicating that this number of repetitions is sufficient for shadowing to effectively improve listening skills (Hamada, 2016a), it is feasible that more repetitions may be more effective. van Paridon et al. (2019) hypothesize that being able to shadow does not necessarily mean that a learner can process the meaning of the input. Thus, for shadowing to effectively improve listening comprehension, the optimal number of repetitions may be greater than four or five times. Indeed, Hashizaki (2024b) showed that more than five repetitions lead to better improvement in listening comprehension through shadowing.

Effect of Repetitions on L2 MWE Learning

The number of repetitions required for L2 MWE learning remains unclear, despite various studies on this topic (Hashizaki, 2021; Lin, 2021; Pellicer-Sánchez, 2017; Peters, 2014; Szudarski & Carter, 2016; Webb et al., 2013). Regarding learning methods other than shadowing, Webb et al. (2013) explored the efficacy of the number of repetitions (1, 5, 10, and 15) on the incidental learning of verb-noun collocations among 161 first- and second-year university students. Their findings revealed a positive correlation between the number of repetitions and collocational learning gains, with 15 exposures yielding the greatest benefit. However, Pellicer-Sánchez (2017) produced contrasting outcomes; their investigation involved 41 L2 learners and focused on the impact of the number of repetitions on the acquisition of adjective pseudowords during reading exercises. Their findings revealed no discernible distinction in terms of incidental collocation learning between repeating the material four or eight times. These conflicting results prompt the need for further research on the influence of the number of repetitions on MWE learning.

As for shadowing, Miyake (2009) demonstrated that approximately six repetitions can improve the speed of repeating phrases and facilitate their subsequent memorization. Hashizaki (2021) investigated whether repeating shadowing up to 30 times could enhance the memorization of multiword expressions (MWEs or chunks). The study involved 20 Japanese EFL learners who performed 30 repetitions of shadowing using two types of materials (easy and difficult). After every 10 repetitions, the participants completed a cued recall test. The results revealed a statistically significant effect of repetition on the memorization of MWEs up to 20 repetitions, provided that the material difficulty was appropriate.

The Present Study

The previous studies have shown that four to five repetitions of shadowing for one material effectively improve L2 listening comprehension (Hamada, 2016a; Tamai, 2005). However, van Paridon et al.'s (2019) model suggests that the ability to shadow a word does not ensure the processing of its meaning, and more repetitions may induce improvements in listening comprehension. Moreover, while shadowing is effective for L2 MWE memorization (Hashizaki, 2021, 2024c; Miyake, 2009; Xing & Hashizaki, 2021, 2024), the optimal number of repetitions remains unestablished. Finally, previous research has not established the appropriate number of repetitions in shadowing training, not only in listening comprehension and the memorization of MWEs but also in their theoretical underpinnings, such as bottom-up processing and repetition speed. Based on these knowledge gaps, this study explores the appropriate number of repetitions in shadowing training by focusing on the following aspects that shadowing aims to improve: listening comprehension, L2 MWE memorization, bottom-up processing, and repetition speed. Accordingly, the research questions (RQ) are as follows:

RQ 1. Can more than five repetitions of shadowing the same passage effectively enhance listening comprehension, the memorization of MWEs, bottom-up processing, and repetition speed?

RQ 2. Do the effects of repeated shadowing differ depending on each of these four aspects?

Method

Participants

This quasi-experimental study utilized one class of 37 nursing majors at a Japanese university. Of the 37 students, 30 provided their consent via a Google Forms issued in the final class of the semester, and their data were used for the analysis. The 30 participants (5 men, 25 women) had an average age of 18.20 years (standard deviation [*SD*] = 0.41). Eight participants did not submit the sorting post-test, so a total of 22 participants completed the sorting test. For the same reason, a total of 26 participants completed the read-aloud test. The vocabulary size test (V_YesNo V1.0; Meara & Miralpeix, 2016) indicated that the participants' vocabulary size was 2,060.93 English words (*SD* = 1,000.17) on average, indicating that they were at the beginner level.

Materials

Shadowing Material

The nursing class adopted the textbook *Medical English Clinic* (Nishihara et al., 2011), which comprised 13 units. Only the first 11 units were used for instruction, concentrating exclusively on each unit's listening section. The publisher provided speech at two rates: normal and slow. Table 1 details the material used.

Table 1

Details of the Medical English Clinic (Nishihara et al., 2011) Textbook Used

Unit	FRE	FKG	Words	WPM (normal)	WPM (slow)
1	91.10	2.00	85.00	117.12	110.80
2	89.60	2.30	91.00	105.94	92.65
3	72.70	5.20	93.00	107.02	93.40
4	99.90	0.70	83.00	121.50	114.02
5	89.80	2.40	90.00	126.20	108.06
6	83.50	3.30	78.00	111.00	106.58
7	97.40	1.60	97.00	129.84	111.96
8	82.30	3.40	144.00	124.76	106.42
9	82.90	3.50	141.00	129.86	104.67
10	96.30	1.50	143.00	134.12	112.68
11	81.80	4.50	186.00	143.80	123.62
M	87.94	2.76	111.91	122.83	107.71
SD	7.78	1.29	33.76	11.24	8.47

Note. FRE stands for Flesch Reading Ease, where a higher score indicates greater readability (Microsoft, 2025). FKG refers to the Flesch–Kincaid Grade Level, with higher values signifying increased difficulty (Microsoft, 2025). "Words" represents the number of words per unit. WPM (normal) indicates the words spoken per minute at a normal pace, while WPM (slow) refers to the words spoken per minute at a slower pace.

Dictation-Cloze Test

To investigate the participants' improvements in bottom-up processing (especially word perception), a dictation-cloze test was conducted at the beginning and end of the semester. The material was extracted from the Voice of America (VOA) (VOA Learning English, 2022). Blanks were made of the VOA material to develop a dictation-cloze test and function words were extracted, such as "particles, prepositions, pro-forms, articles, be verbs, auxiliary verbs, and conjunctions that carry relational meaning rather than lexical meaning" (Rost, 2015, p. 286). Only function words were targeted for the dictation-cloze test because the ability to dictate content words is susceptible to learners' vocabulary knowledge, which makes it difficult to measure improvements in pure bottom-up processing (Hamada, 2016a). The test was conducted using paper and pencil, but the participants sent their answers through Google Forms. In the forms, the participants entered each written word in the blanks. No explanation was provided after the pre-test to avoid the retest effect.

Listening Test

This study adopted a standardized English proficiency test in Japan known as *Eiken*. This test's levels are divided into seven categories, starting from the easiest: grades 5, 4, 3, Pre-2, 2, Pre-1, and 1 (Eiken Foundation of Japan, 2023). Thirty listening comprehension questions from the Eiken Grade 3 test conducted in January 2022 were used. The first author conducted the test using Google Forms before and after the learning sessions. The same test was conducted twice (once before 11 training sessions, and once after), but there were no explanations of the content of the pre-test to avoid students learning the test contents. By taking the same test multiple times, there is a possibility that the score on the second test may be higher due to the test-retest effect. However, since the current study demonstrates the effect of repeated shadowing, this is not considered a critical issue (refer to the results section).

Sorting Test

This study employed a sorting test in which the participants placed words in the correct order to create the correct MWEs. The sorting test involved 33 expressions, each consisting of multiple words, taken from the *Medical English Clinic* (Nishihara et al., 2011) textbook. Three MWEs were selected from each textbook unit. This assessment was conducted at the beginning and end of the semester. Additionally, three MWEs from each unit were individually tested at the end of each class, although the data were not included in the analysis. This is because the data were not necessary to answer the research questions in this study.

Read-Aloud Test

The read-aloud test was conducted before and after the learning sessions. Recordings were made using the students' smartphones. In the pre-test, two materials were read aloud: one served as the training material (the listening section of Unit 6 from *Medical English Clinic* textbook), and the other was "text A," employed from Saito and Saito (2017) but not included in the training (Table 2). In addition to the two pre-test materials, the post-test also used "text B" from Saito and Saito (2017), although it was not included in the analysis because "text B" was for another study aimed at examining improvements in pronunciation, which is beyond the scope of the present study.

Table 2
Materials Used for the Read-Aloud Test

Material	FRE	FKG	Words	WPM (normal)	WPM (slow)
Learned item (Dialogue 6 [Unit 6] from the study material)	83.50	3.30	78.00	111.00	106.58
Control item (text A from Saito and Saito [2017])	73.10	5.60	53.00		

Note. Refer to Table 1 for the explanations of the terms. The WPM for the control item was not available because this item was used only in the read-aloud test.

Procedure

The research was conducted over 14 classes. This study utilized the first two classes and the last class to assess the participants' vocabulary size, listening comprehension, MWE knowledge, bottom-up processing, and repetition speed, while the remaining 11 classes were used for the shadowing training. Before the tests, students were informed that the results would not impact their grades and were solely intended to assess their English proficiency at that time. Therefore, although not impossible, it was deemed unlikely that students would attempt to cheat.

In the first class, the read-aloud and vocabulary size tests were conducted. For the read-aloud test, participants read the first material aloud and recorded their speech with cues from the first author. Then, they repeated the procedure for the second material. After recording, they sent the recorded file to the first author through Google Forms. The vocabulary size test was explained in class, and participants completed it as homework at their convenience. The first author noted that the test could detect if participants were not fully engaged, such as by answering without consideration. In the second class, the listening and dictation-cloze tests were conducted. The first author administered the listening test via a classroom speaker, and the responses were collected using Google Forms. The dictation-cloze test was conducted using pencil and paper, with the first author playing the target speech through the speaker. To prevent any test-related effects, no explanations were provided for the two tests. Moreover, an MWE knowledge test was provided as homework via Google Forms. The participants completed this test themselves.

The shadowing training comprised three stages. In the first stage, the first author presented an explanation of the material content and described the vocabulary, grammar, syntax, and meaning of the dialogue. In the second stage, pronunciation, encompassing both segmental and suprasegmental features, was explained for half of the material. Subsequently, the participants individually identified the segmental and suprasegmental features in the remaining material and shared their findings in pairs or small groups. In the last stage, the participants engaged in shadowing training for 10–15 minutes, during which they could refer to the textbook and seek clarification from the first author. They conducted the training individually, using their smartphones and headphones. When headphones were unavailable, they placed the phone speaker close to their ear to listen to the material. After completing the shadowing training, the participants completed sorting and dictation-cloze tests on the material of the day. After class, the participants were instructed to shadow the material of the day at least 10 times as homework. In addition, they submitted their recordings for the 10 repetitions to confirm completion. In the 14th class, the participants underwent the same dictation-cloze, listening, and read-aloud tests. The sorting test was conducted as homework. Refer to Table 3 for the study overview.

Table 3
Study Overview

Day	Test
1	• Vocabulary size test (homework) • Read-aloud test
2	• Listening test • Dictation-cloze test • Sorting test (homework)
3–13	• Shadowing training
14	• Listening test • Dictation-cloze test • Read-aloud test • Sorting test (homework)

Note. Day 15 was used for giving feedback to participants based on the test results.

Analysis

The study employed generalized linear mixed-effect modeling (GLMM) using R 3.5.3 (R Core Team, 2022) and the *lme4* package (Bates et al., 2015). For the listening, sorting, and dictation-cloze tests, binomial distribution and logit link function were applied. For the read-aloud test, gaussian distribution and an identity link function were applied. To prevent convergence errors, the categorical variables were simple-coded and the numerical variables were normalized. The fixed effects in the GLMM included timing (pre vs. post) and the number of repetitions (NoR: the total repetitions inside and outside the classroom). Condition (learned vs. control) was included only for the read-aloud test. The response variables encompassed accuracy on the listening, sorting, and dictation-cloze tests (scored 1 for correct answers and 0 for incorrect answers). For the read-aloud test, Words Per Minute (WPM) was calculated based on the duration of the read-aloud test, and this was a continuous variable. The random intercepts comprised the participants and items, with random slopes incorporated solely for the within-participant and within-item conditions in alignment with the study's design rationale (Barr et al., 2013).

The model selection followed a systematic approach. Initially, the maximal model incorporated all fixed effects, their interactions, and the justified random slopes. Subsequently, the insignificant fixed effects and interactions were removed, and the random slopes that did not enhance the model fit were excluded to prevent Type II error (Matuschek et al., 2017). For the model comparison, the Akaike information criterion (AIC) scores were used; a lower AIC score indicated a better model fit. The *anova* function was used to assess whether an extracted fixed effect or interaction contributed to improved model fit. This iterative process was continued until no more random slopes were identified that enhanced the model fit. When an interaction was statistically significant, the simple main effects were examined using the *phia* package (De Rosario-Martinez et al., 2023).

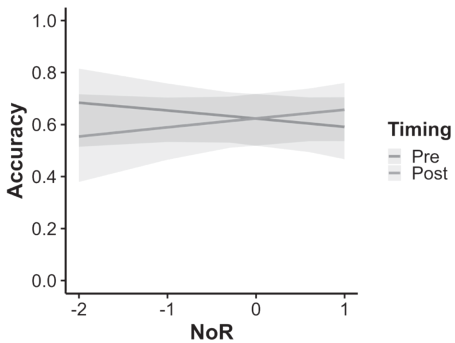
Results

To reveal whether the total required NoR differed depending on the shadowing effects, four individual models were created for the listening test (listening comprehension), sorting test (memorization of MWEs), dictation-cloze test (bottom-up processing), and read-aloud test (repetition speed).

The final model for the listening test included the main effects of timing (Estimate = 0.003, *SE* = 0.106, *z* = 0.024, *p* = .981), NoR (Estimate = 0.005, *SE* = 0.138, *z* = 0.034, *p* = .973), and their interaction (Estimate = 0.278, *SE* = 0.105, *z* = 2.641, *p* = .008). The random intercepts were included for the participants and items. No random slopes were included because they did not significantly improve the AIC score. The results demonstrated higher pre-test accuracy for participants with less NoR, and higher post-test accuracy for participants with higher NoR (Figure 2). Table 4 presents the model details.

Figure 2

Interaction Between Timing and NoR in the Listening Test



Note. Timing refers to the time at which the participants took the test. NoR refers to the number of repetitions.

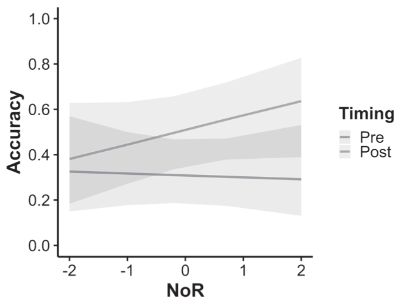
Table 4
GLMM Results for the Listening Test

	Fixed effect				Random effect	
					Participant	Item
Parameter	Estimate	*SE*	*z*	*p*	*SD*	*SD*
(Intercept)	0.504	0.203	2.479	$p = .013$	0.701	0.813
Timing: post-test	0.003	0.106	0.024	$p = .981$	-	-
NoR	0.005	0.138	0.034	$p = .973$	-	-
Timing: post-test * NoR	0.278	0.105	2.641	$p = .008$	-	-

Note. Number of observations = 1,800; $n = 30$. Model formula: Accuracy ~ Timing * NoR + (1|Participant) + (1|Item). Timing refers to the time at which the participants took the test; NoR refers to the number of repetitions. Marginal $R^2 = 0.004$, conditional $R^2 = 0.262$. Timing was simple-coded (Pre = −0.5; Post = 0.5).

The model for the sorting test included the main effects of timing (Estimate = 0.845, $SE = 0.133$, $z = 6.343$, $p < .001$), NoR (Estimate = 0.111, $SE = 0.182$, $z = 0.606$, $p = .545$), and their interaction (Estimate = 0.300, $SE = 0.129$, $z = 2.327$, $p = .020$). The random intercepts were included for participants and items; however, no random slopes were used because they did not significantly improve the AIC score. Table 5 presents the model details. In the pre-test, NoR did not appear to affect accuracy. However, in the post-test, learners with high NoR tended to achieve better accuracy (Figure 3).

Figure 3
Interaction Between Timing and NoR in the Sorting Test



Note. Timing refers to the time at which the participants took the test. NoR refers to the number of repetitions.

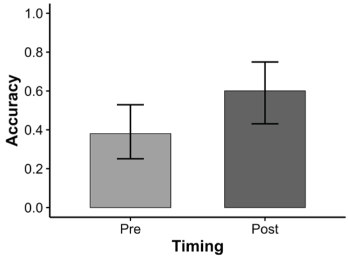
Table 5
GLMM Results for the Sorting Test

	Fixed effect				Random effect	
					Participant	Item
Parameter	Estimate	*SE*	*z*	*p*	*SD*	*SD*
(Intercept)	−0.385	0.334	−1.155	p = .248	0.800	1.593
Timing: post-test	0.845	0.133	6.343	p = .001	-	-
NoR	0.111	0.182	0.606	p = .545	-	-
Timing: post-test * NoR	0.301	0.129	2.327	p = .020	-	-

Note. Number of observations = 1,452; n = 22. The number of participants was smaller than that in the other two tests because eight students did not submit the sorting post-test. Model formula: Accuracy ~ Timing * NoR + (1|Participant) + (1|Item). Timing refers to the time at which the participants took the test; NoR refers to the number of repetitions. Marginal R^2 = 0.032, conditional R^2 = 0.508. Timing was simple-coded (Pre = −0.5; Post = 0.5).

The final model for bottom-up processing included the main effect of timing (Estimate = 0.900, SE = 0.171, z = 5.241, p < .001). The random intercepts were included for participant and items, and timing was included as a random slope for the items. Thus, the post-test accuracy exceeded that of the pre-test, irrespective of the NoR (Figure 4). Table 6 presents the model details.

Figure 4
Pre- and Post-Test Scores for the Dictation Test



Note. Timing refers to the time at which the participants took the test.

Table 6
GLMM Results for the Dictation Test

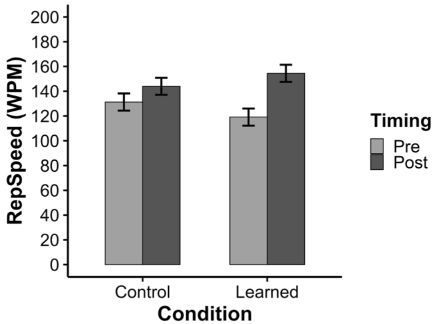
	Fixed effect				Random effect	
					Participant	Item
Parameter	Estimate	*SE*	*z*	*p*	*SD*	*SD*
(Intercept)	−0.040	0.319	−0.125	p = .900	0.854	1.348
Timing: post-test	0.897	0.171	5.241	p = .001	-	-

Note. Number of observations = 1,500; n = 30. Model formula: Accuracy ~ Timing + (1 | Participant) + (1 + Timing | Item). Timing refers to the time at which the participants took the test. Marginal R^2 = 0.033, conditional R^2 = 0.462. Timing was simple-coded (Pre = −0.5; Post = 0.5).

The final model for repetition speed included the main effects of timing (Estimate = 24.040, SE = 2.667, t = 9.015, p< .001) and condition (Estimate = −0.828, SE = 2.667, t= −0.310, p = .757). The interaction between timing and condition was also included in the model (Estimate = 22.636, SE = 5.333, t = 4.244, p < .001). The random intercepts were included for participants and items, and timing was included as a random slope for participants. Since the interaction between timing and condition was significant, a simple effect test was conducted. The results showed that the effect of timing was significant in both conditions (Learned and Control). This meant that repetition speed improved in both conditions, although that of the learned item improved to a larger degree (Figure 5). The effect of condition was also significant in both timings (Pre- and Post-Tests). The results showed that the WPM of the learned item was lower than that of the controlled item in the pre-test, while the WPM of the learned item was significantly higher than that of the controlled item in the post-test. The NoR was not included in the model because it was not significant in any models. Table 7 presents the model details.

Figure 5

Interaction Between Timing and Condition for Repetition Speed (WPM)



Note. Timing refers to the time at which the participants took the test. Condition represents whether the items were learned in the training sessions (Learned) or not (Control).

Table 7

LME Results for the Read-Aloud Test

					Random effect
	Fixed effect				Participant
Parameter	Estimate	*SE*	*t*	p	*SD*
(Intercept)	137.227	2.607	52.645	$p < .001$	11.420
Timing: post-test	24.040	2.667	9.015	$p < .001$	-
Condition: learned	−0.828	2.667	−0.310	$p = .757$	-
Timing: post-test * Condition: learned	22.636	5.334	4.244	$p < .001$	-

Note. Number of observations = 104; n = 26. Model formula: WPM ~ Timing * Condition + (1 | Participant). Timing refers to the time at which the participants took the test. Marginal R^2 = 0.365, conditional R^2 = 0.631. Timing was simple-coded (Pre = −0.5; Post = 0.5).

Discussion

This study investigated whether repeated shadowing (more than five times) could effectively improve L2 listening comprehension, the memorization of MWEs, and their theoretical underpinnings: bottom-up processing and repetition speed (RQ 1), and whether the required number of repetitions varied based on these aspects of shadowing effects (RQ 2). Overall, the results showed that, for listening comprehension and MWEs memorization, more repetitions were required for post-test improvements while bottom-up processing and repetition speed improved irrespective of the number of repetitions.

Effects of Repetitions on the Four Tests

Regarding the participants' listening comprehension, more repetition appeared to be important for shadowing to be effective. This is in line with Hashizaki (2024b), who showed that the more the learners repeat shadowing, the better the improvement of listening comprehension of the learners. This result can be explained via van Paridon et al.'s (2019) model, which asserts that there are two shadowing pathways: one in which meaning is processed and one in which a learner imitates sounds immediately after the segmentation or selection of phonological codes without engaging in processing meaning. Thus, although O'ki (2014) and Shiki et al. (2010) have shown that shadowing repetition of four to five times leads to a plateau in the reproduction rate of shadowing, this does not necessarily mean that the process becomes automatized and leads to the processing of meaning, as van Paridon et al.'s (2019) model suggests. In the current study, several repetitions might have been sufficient to improve the participants' ability to perceive words in speech (as measured by the dictation-cloze test) and repeat perceived words quickly (as measured by the read-aloud test); however, this might have been insufficient for automatizing participants' bottom-up processing and improving their general listening comprehension. Thus, more repetitions seem to be a prerequisite for shadowing to effectively improve listening comprehension.

Regarding the memorization of MWEs, the more the participants repeated the shadowing, the more MWEs they memorized. This finding corroborates that of previous research, which has shown that repetition is required when learning MWEs (Hashizaki, 2021; Lin, 2021; Pellicer-Sánchez, 2017; Peters, 2014; Szudarski & Carter, 2016; Webb et al., 2013). Although the exact number of required repetitions remains unclear, encouraging more than five repetitions is recommended for shadowing to effectively aid in the memorization of MWEs.

Regarding bottom-up processing and repetition speed, increased NoR was not necessary for significant post-test improvements. For the dictation-cloze test, this may have been because this test specifically assessed the recognition of function words; this ability did not require automaticity because the participants were given sufficient time to write down the target words. Concerning repetition speed, the participants' ability to read words aloud might have been achieved through processing the sounds rather than the meanings of the words. This finding is in line with that of van Paridon et al. (2019), who state that the processing of shadowing has two routes: one which processes sounds and one which processes both sounds and meanings. This study's dictation-cloze test and repetition speed findings also agree with those of the previous shadowing studies that have found that the optimal number of repetitions is four to five (Hamada, 2016a; O'ki, 2014; Shiki et al., 2010). Shiki et al. (2010) state that four to five repetitions are sufficient for shadowing to be effective because this leads to a plateau in the reproduction rate of shadowing. Hamada (2016a) suggests that six repetitions per material are sufficient to improve bottom-up processing (word perception), irrespective of the participants' proficiency. Thus, more than five repetitions do not seem necessary to achieve improvements in bottom-up processing through shadowing. Similarly, Tamai (2005) states that the accuracy of repeating words (which requires word recognition) and repetition speed are achieved in the early stage of training as shown in Kadota's (2019) model (Figure 1).

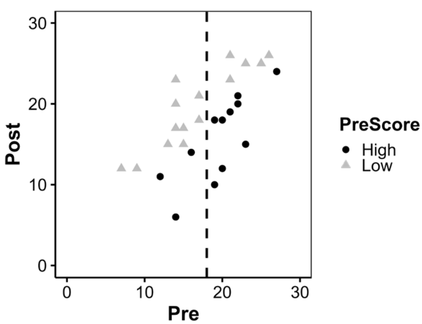
Explaining the Discrepancies Between the Past Research and Current Study

Regarding the listening test, this study's results contrast with those of Hamada (2016a) and Tamai (2005), who found that approximately four to five shadowing repetitions could improve listening comprehension for university students with low- and intermediate-proficiency levels. This can be potentially explained in terms of data analysis methods. Specifically, dividing learners into proficiency groups based on their pre-test scores and treating them as categorical variables might have favored the detection of the effect of shadowing on listening comprehension. Hamada (2016a) and Tamai (2005) used their participants' pre-test scores to indicate proficiency and divided them into two and three groups as categorical variables, respectively. They then analyzed the interaction between the proficiency levels (low vs. intermediate for Hamada [2016a]; low vs. mid vs. high for Tamai [2005]) and test timings (pre vs. post). Their results showed that the interactions were statistically significant; low-proficiency learners in Hamada's (2016a) study and low- and mid-proficiency learners in Tamai's (2005) study showed improved post-test listening comprehension. However, this analysis method might have overestimated the effect of shadowing by selecting data points with the potential for improvement. It also excludes learners whose scores decreased between the pre- and post-tests. This is supported by the regression toward the mean, which is "a phenomenon [wherein] a variable that is extreme on its first measurement will tend to be closer to the center of the distribution in a later measurement" (Everitt & Skrondal, 2010, pp. 363–364). Based on this phenomenon, the high scores in the pre-test tend to become low in the post-test, and the low scores in the pre-test tend to become high in the post-test based on the median of the pre-test. Thus, it is plausible that the low scores will become higher in the post-test due to this statistical phenomenon. This can ultimately lead to Type I error (Kusanagi & Tamura, 2017). Figure 6 shows a scatterplot of this study's pre- and post-test scores. Indeed, Hashizaki (in press) indicated the possibility that considering pre-test scores as a measure of proficiency may lead to an overestimated effect of shadowing on low-proficiency learners

In this figure, the dotted line represents the median value of the pre-test scores. The left side of the line signifies a "low" score while the right side indicates a "high" score. Among the participants who score below the median on the pre-test, only three show a decrease in their post-test scores. Conversely, among the participants who score above the median on the pre-test, 12 exhibit lower post-test scores compared to their pre-test scores.

Figure 6

Scatterplot of the Pre- and Post-Test Scores for the Listening Test



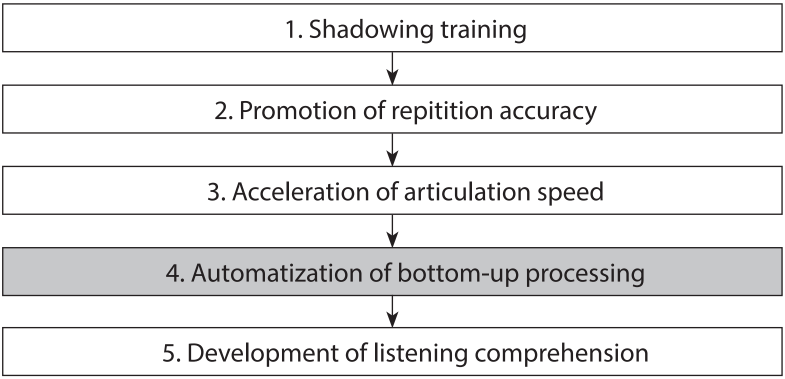
Note. The number of dots totaled 27 (n = 30) because there were three data points that had the same pre- and post-test scores.

A Possible Model to Explain the Study Results

The present study found that extensive repetition was not necessary to improve the participants' bottom-up processing and repetition speed. On the other hand, more repetitions were necessary for enhancing listening comprehension. This may suggest that bottom-up processing needs to become automated in order to free up cognitive resources for meaning processing, which in turn facilitates listening comprehension (Figure 7).

Figure 7

A Possible Revision of the Mechanism Through Which Shadowing Improves Listening Comprehension



Implications of the Findings

The results indicated two pedagogical and methodological implications for the teaching and studying of shadowing. From a pedagogical perspective, shadowing appears to require more repetition than previously thought to effectively improve listening comprehension. While it is not possible to specify the exact number of required repetitions, a general guideline is that learners should repeat shadowing until they can do so automatically while paying attention to the meaning of the material.

Methodologicaly, this study discusses the potential for overestimating the effects of shadowing owing to the analysis method; this applies not only to shadowing but also to the effectiveness of other learning methods. Therefore, to accurately measure effects in real-world settings in the future, as Hashizaki (in press) suggests, participants should not be divided into subgroups based on their pre-test scores, as this may lead to Type I errors. Alternatively, when examining the impact of proficiency levels, it is advisable to assess English proficiency separately, use the score to define proficiency levels, and then analyze the interaction between proficiency and test timing.

Limitations and Future Study

With all the findings, this study had the following limitations. First, while it argued that the number of repetitions affects automatization in the model, the actual automatization process was not examined. To assess the speed of bottom-up processing, tasks involving the judgment of phrases or sentences using audio should be employed in the future. Second, although repetition may facilitate meaning processing during shadowing, this study did not include any questionnaires or tests to measure this. Therefore, future studies should employ measurements that enable the observation of improvements in the processing of meaning during shadowing, such as questionnaires on shadowing strategies or interpretation tests of shadowed materials. Third, regarding the idea that fewer repetitions are effective owing to the consideration of pre-test scores as proficiency indicators, a separate proficiency test should be conducted in the future to confirm this assertion. Tests measuring actual proficiency levels should be conducted to provide clearer insights into whether the prior studies' effectiveness of using fewer repetitions was due to their consideration of pre-test scores as proficiency indicators. Fourth, although this study suggests that more repetitions are needed for improving listening comprehension and memorizing MWEs through shadowing, the specific number of repetitions required was not thoroughly investigated. Therefore, future research should employ statistical methods to clarify the effectiveness of repetitions up to a certain number. Achieving these objectives can help to elucidate the process of improving listening ability through shadowing and delineate its effectiveness. Fifth, this study did not establish a control group. While the number of repetitions significantly influenced the results, suggesting that shadowing was effective, future studies should include a control group to exclude the possibility that factors other than shadowing influenced the improvement of the measured skills. Finally, this study focused solely on the effects of shadowing on listening and MWE memory. However, some research suggests that shadowing also impacts speaking abilities, such as fluency (Muraoka, 2019) and pronunciation (Foote & McDonough, 2017; Niimoto, 2022). Future research could benefit from examining the effects of repetition on the improvement of these abilities using a speaking test.

Conclusion

This study examined whether the effects of repetition varied based on different aspects of shadowing effectiveness. First, to automatize bottom-up processing and observe improved listening comprehension, more than five repetitions may be necessary. Contrarily, enhanced bottom-up processing and faster repetition rate could be observed with fewer repetitions. Second, a higher number of repetitions was essential for the retention of MWEs in a learner's memory. Third, the consideration of a learner's pre-test score as a proficiency indicator could suggest an overestimation of shadowing effectiveness. Finally, this study proposed a model to explain the results, incorporating the additional component of the automatization of bottom-up processing.

Acknowledgments

This research was funded by JST SPRING (Grant Number: JPMJSP2125). The first author gratefully acknowledges the support of the "Interdisciplinary Frontier Next-Generation Researcher Program of the Tokai Higher Education and Research System." Sincere thanks are also extended to the Hattori International Scholarship Foundation for their generous financial assistance.

Ryotaro Hashizaki is an assistant professor at the Faculty of International Studies, University of Economics and Law. His research interests lie in listening and formulaic language, with a particular focus on the effectiveness of shadowing and oral reading.

Hirohisa Sekiyama is a Ph.D. student at the Graduate School of Language, Communication, and Culture, Kwansei Gakuin University. His interests fall under the umbrella of second-language listening instruction, such as shadowing and dictogloss.

Yun Xing is a lecturer at the School of Foreign Languages, Shanxi University. She is interested in EFL learners' acquisition of multiword units. Her recent publications include the "Effects of repetition, word types, and material difficulty on the reproduction rate of shadowing" published in the LET Journal of Central Japan.

Kumi Wakita is a Master's student at the Graduate School of Humanities, Nagoya University. She is interested in the cognitive mechanisms involved in written and spoken language processing. Her particular interest lies in the processing of lexical bundles by EFL learners.

References

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *14*(6), 575–589. https://doi.org/10.1016/s0022-5371(75)80045-4

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Conklin, K., & Schmitt, N. (2012). The processing of formulaic language. *Annual Review of Applied Linguistics*, *32*, 45–61. https://doi.org/10.1017/S0267190512000074

De Rosario-Martinez, H., Fox, J., R Core Team, & Chalmers, P. (2023). *Package "phia."* https://github.com/heliosdrm/phiahttps://cran.r-project.org/web/packages/phia/index.html

Eiken Foundation of Japan. (2023). *EIKEN Grades*. https://www.eiken.or.jp/eiken/en/grades/

Everitt, B. S., & Skrondal, A. (2010). *The Cambridge dictionary of statistics*. Cambridge University Press. https://www.cambridge.org/jp/universitypress/subjects/statistics-probability/statistics-and-probability-general-interest/cambridge-dictionary-statistics-4th-edition?format=HB&isbn=9780521766999

Fawcett, J. M., & Ozubko, J. D. (2016). Familiarity, but not recollection, supports the between-subject production effect in recognition memory. *Canadian Journal of Experimental Psychology*, *70*(2), 99–115. https://doi.org/10.1037/cep0000089

Field, J. (2003). Promoting perception: Lexical segmentation in L2 listening. *ELT Journal*, *57*(4), 325–334. https://doi.org/10.1093/elt/57.4.325

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

Hamada, Y. (2016a). Shadowing: Who benefits and how? Uncovering a booming EFL teaching technique for listening comprehension. *Language Teaching Research*, *20*(1), 35–52. https://doi.org/10.1177/1362168815597504

Hamada, Y. (2016b). *Teaching EFL learners shadowing for listening: Developing learners' bottom-up skills*. Routledge. https://doi.org/10.4324/9781315677118

Hashizaki, R. (2021). The effect of shadowing on memorization of English chunks and its theoretical basis: Influence of repetition and material difficulty. *LET Journal of Central Japan*, *31*, 39–52. https://doi.org/10.20656/letcj.31.0_39

Hashizaki, R. (2024a). Does vocalization enhance L2 formulaic sequence learning? *Language Teaching Research*. https://doi.org/10.1177/13621688231224040

Hashizaki, R. (2024b). Repetition fosters the effect of shadowing on L2 listening skills: Seeking an enhanced procedure. *The Journal of Asia TEFL*, *21*(3), 600–617. https://doi.org/10.18823/asiatefl.2024.21.3.6.600

Hashizaki, R. (2024c). Theoretical validation of the effect of shadowing on the memorization of multiword units. *JACET Kansai Journal*, *26*, 1–19.

Hashizaki, R. (in press). Re-examining the effect of shadowing: Is it the method for novice L2 learners? *JACET Journal, 69*.

Icht, M., & Mama, Y. (2022). The effect of vocal production on vocabulary learning in a second language. *Language Teaching Research*, *26*(1), 79–98. https://doi.org/10.1177/1362168819883894

Kadota, S. (2015). *Shadowing, ondoku to eigo komyunikeisyon no kagaku* [Science of shadowing, oral reading, and English communication]. Cosmopier Publishing Company. https://www.cosmopier.com/shoseki/4864540756/

Kadota, S. (2019). *Shadowing as a practice in second language acquisition: Connecting inputs and outputs*. Routledge. https://doi.org/10.4324/9781351049108-1

Kusanagi, K., & Tamura, Y. (2017). 外国語教育研究における事後分析の危険性.[The risks of post-hoc analyses in foreign language education research] The Japan Association for Language Education & Technology Chubu Branch, Foreign Language Education Foundation Research Subcommittee, Annual Report Collection for the Year 2016. https://www.letchubu.net/modules/xpwiki/?2016%E5%B9%B4%E5%BA%A6%E5%A0%B1%E5%91%8A%E8%AB%96%E9%9B%86

Lin, P. (2021). In search of the optimal mode of input for the acquisition of formulaic expressions. *TESOL Quarterly*, *55*(3), 1011–1023. https://doi.org/10.1002/tesq.3055

MacLeod, C. M., Gopie, N., Hourihan, K. L., Neary, K. R., & Ozubko, J. D. (2010). The production effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(3), 671–685. https://doi.org/10.1037/a0018785

Matuschek, H., Kliegl, R., Vasishtha, S., Baayen, H., & Bates, D. (2017). Balancing type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315. https://doi.org/10.1016/j.jml.2017.01.001

Meara, P., & Miralpeix, I. (2016). *Tools for researching vocabulary*. Multilingual Matters. https://doi.org/10.21832/9781783096473

Microsoft. (2025). *Get your document's readability and level statistics*. Microsoft Support. https://support.microsoft.com/en-us/office/get-your-document-s-readability-and-level-statistics-85b4969e-e80a-4777-8dd3-f7fc3c8b3fd2

Miyake, S. (2009). Cognitive processes in phrase shadowing: Focusing on articulation rate and shadowing latency. *JACET Journal*, *48*, 15–28. https://iss.ndl.go.jp/books/R000000004-I10222258-00

Muraoka, Y. (2019). The effect of shadowing training on the development of second language fluency. *Seigakuin Univeristy Bulletin*, *66*(2), 46–49. https://serve.repo.nii.ac.jp/record/3833/files/sk66_(46).pdf

Niimoto, S. (2022). Shadowing with reading-aloud training for better comprehensibility. *Studies in English Language Teaching*, *45*, 145–154. https://doi.org/10.18989/selt.45.0_145

Nishihara, T., Nishihara, M., & Brown, T. (2011). *Medical English clinic*. Cengage Learning. https://cengagejapan.com/elt/ESP/page/?no=1256714766fa6gd

O'ki, T. (2011). Classification of learners' repetition strategies in shadowing at the beginning stage. *KATE Journal*, *25*, 33–43. https://doi.org/10.20806/katejournal.25.0_33

O'ki, T. (2014). Effectiveness of repetition on shadowing: Can learners repeat more words? *Hakuoh University Journal*, *28*(2), 169–187. https://hakuoh.repo.nii.ac.jp/?action=pages_view_main&active_action=repository_view_main_item_detail&item_id=1599&item_no=1&page_id=13&block_id=21

Ozubko, J. D., Hourihan, K. L., & MacLeod, C. M. (2012). Production benefits learning: The production effect endures and improves memory for text. *Memory*, *20*(7), 717–727. https://doi.org/10.1080/09658211.2012.699070

Pellicer-Sánchez, A. (2017). Learning L2 collocations incidentally from reading. *Language Teaching Research*, *21*(3), 381–402. https://doi.org/10.1177/1362168815618428

Peters, E. (2014). The effects of repetition and time of post-test administration on EFL learners' form recall of single words and collocations. *Language Teaching Research*, *18*(1), 75–94. https://doi.org/10.1177/1362168813505384

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rost, M. (2015). *Teaching and researching listening*. Routledge. https://doi.org/10.4324/9781315732862

Saito, Y., & Saito, K. (2017). Differential effects of instruction on the development of second language comprehensibility, word stress, rhythm, and intonation: The case of inexperienced Japanese EFL learners. *Language Teaching Research*, *21*(5), 589–608. https://doi.org/10.1177/1362168816643111

Shiki, O., Mori, Y., Kadota, S., & Yoshida, S. (2010). Exploring differences between shadowing and reproduction practices: An analysis of reproduction rate and types of reproduced words. *ARELE: Annual Review of English Language Education in Japan*, *21*, 81–90. https://doi.org/10.20581/arele.21.0_81

Suzuki, Y. (2023). *Practice and automatization in second language research* (1st ed.). Routledge. https://doi.org/10.4324/9781003414643-1

Szudarski, P., & Carter, R. (2016). The role of input flood and input enhancement in EFL learners' acquisition of collocations. *Vigo International Journal of Applied Linguistics*, *26*(2), 245–265. https://doi.org/10.1111/ijal.12092

Tamai, K. (2005). リスニング指導法としてのシャドーイングの効果に関する研究. [Research on the effect of shadowing as a listening instruction method] Kazamashobo.

Tang, J. (2013). Input of chunks and its effects on L2 learners' listening competency. *Theory and Practice in Language Studies*, *3*(7), 1264–1269. https://doi.org/10.4304/tpls.3.7.1264-1269

Vandergrift, L. (2004). Listening to learn or learning to listen? *Annual Review of Applied Linguistics*, *24*, 3–25. https://doi.org/10.1017/S0267190504000017

van Paridon, J., Roelofs, A., & Meyer, A. S. (2019). A lexical bottleneck in shadowing and translating of narratives. *Language, Cognition and Neuroscience*, *34*(6), 803–812. https://doi.org/10.1080/23273798.2019.1591470

Vandergrift, L. (2007). Recent developments in second and foreign language listening comprehension research. *Language Teaching*, *40*(3), 191–210. https://doi.org/10.1017/s0261444807004338

VOA Learning English. (2022). *Italian man lives without electricity*. VOA—Voice of America English News. https://learningenglish.voanews.com/a/italian-man-lives-without-electricity/6887570.html

Webb, S., Newton, J., & Chang, A. (2013). Incidental learning of collocation. *Language Learning*, *63*(1), 91–120. https://doi.org/10.1111/j.1467-9922.2012.00729.x

Xing, Y., & Hashizaki, R. (2021). The effect of shadowing on the acquisition of multiword units using materials at different degrees of difficulty: Focusing on Chinese English learners. *Journal of the Chubu English Language Education Society*, *50*, 73–80. https://doi.org/10.20713/celes.50.0_73

Xing, Y., & Hashizaki, R. (2024). The memorization of multiword units on shadowing tasks: Focusing on learner proficiency, material difficulty, and item length. *Language Education & Technology*, *61*, 1–27. https://doi.org/10.24539/let.61.0_1