The Language Teacher

jalt journal

The research journal of the Japan Association for Language Teaching

Volume 47 • No. 1 • May 2025

JALT
全国語学教育学会
Japan Association for Language Teaching
¥1,900 ISSN 0287-2420

JALT Journal

Volume 47 • No. 1
May 2025

Contents

May 2025
Volume 47 • No. 1

Japan Association for Language Teaching

A Nonprofit Organization

The Japan Association for Language Teaching (JALT) is a nonprofit, professional organization dedicated to the improvement of language teaching and learning in Japan. It provides a forum for the exchange of new ideas and techniques and offers a means of keeping informed about developments in the rapidly changing field of second and foreign language education. Established in 1976, JALT serves an international membership of approximately 3,000 language teachers. JALT has 32 JALT chapters and 32 special interest groups (SIGs) and is a founder of PAC (Pan-Asian Consortium), which is an association of language teacher organizations in Pacific Asia. PAC holds annual regional conferences and exchanges information among its member organizations. JALT is the Japan affiliate of International TESOL (Teachers of English to Speakers of Other Languages) and a branch of IATEFL (International Association of Teachers of English as a Foreign Language). JALT is also affiliated with many other international and domestic organizations.

JALT publishes *JALT Journal*, a semiannual research journal; *The Language Teacher*, a bimonthly periodical containing articles, teaching activities, reviews, and announcements about professional concerns; and the annual *JALT Postconference Publication*.

The JALT International Conference on Language Teaching and Learning and Educational Materials Exposition attracts some 2,000 participants annually and offers more than 600 papers, workshops, colloquia, and poster sessions. Each JALT chapter holds local meetings. and JALT's SIGs provide information and newsletters on specific areas of interest. JALT also sponsors special events such as workshops and conferences on specific themes and awards annual grants for research projects related to language teaching and learning.

Membership is open to those interested in language education and includes copies of JALT publications, free or discounted admission to JALT-sponsored events, and optional membership in one chapter and one SIG. For an annual fee of ¥2,000 per SIG, JALT members can join as many additional SIGs as they desire. For information about JALT membership, contact the JALT Central Office or visit the JALT website.

JALT National Officers (2025)

President: . . . . . *Clare Kaneko*
Vice President: . . . . . *Kenn Gale*
Auditor: . . . . . *Robert Chartrand*
Director of Treasury: . . . . . *Michael Mielke*
Director of Records: . . . . . *Samantha Kawakami*
Director of Program: . . . . . *Chelanna White*
Director of Conference: . . . . . *Wayne Malcolm*
Director of Membership: . . . . . *Julie Kimura*
Director of Public Relations: . . . . . *William Pellowe*

Chapters

Akita, East Shikoku, Fukuoka, Gunma, Hiroshima, Hokkaido, Hokuriku, Ibaraki, Iwate-Aomori, Kitakyushu, Kobe, Kyoto, Matsuyama, Nagano, Nagoya & Gifu, NanKyu, Nara, Niigata, Oita, Okayama, Okinawa, Osaka, Saitama, Sendai, Shizuoka, Tochigi, Tokyo, Tottori, Toyohashi, Yamagata, Yokohama

Special Interest Groups

Accessibility in Language Learning; Art, Research, and Teaching; Bilingualism; Business Communication; CEFR and Language Portfolio; College and University Educators; Computer Assisted Language Learning; Critical Thinking; Extensive Reading; Gender Awareness in Language Education; Global Englishes; Global Issues in Language Education; Intercultural Communication in Language Education; Learner Development; Lifelong Language Learning; Listening; Literature in Language Teaching; Materials Writers; Mind, Brain, and Education; Mixed, Augmented, and Virtual Realities; Other Language Educators; Performance in Education; Pragmatics; School Owners; Study Abroad; Task-Based Learning; Teacher Development; Teachers Helping Teachers; Teaching Younger Learners; Testing and Evaluation; Vocabulary

JALT Central Office

Level 20, Marunouchi Trust Tower–Main,
1-8-3 Marunouchi, Chiyoda-ku, Tokyo 100-0005 Japan
Tel.: (+81) 3-5288-5443; Email: jco@jalt.org;
Website: <https://jalt.org>

In This Issue

This issue contains three full-length research articles, one *Research Forum* article, and three book reviews.

Articles

The first full-length article by **Ryotaro Hashizaki**, **Hirohisa Sekiyama**, **Yun Xing**, and **Kumi Wakita** offers a study that involves 30 Japanese university students taking part in 11 shadowing training sessions, each lasting 90 minutes per week. The effects of these interventions are explored through pre- and post-intervention testing. The authors pay specific attention to how shadowing enhances second-language learners' listening comprehension and memorization of multiword expressions. Specific areas of interest in their study include listening comprehension, multiword expression knowledge, bottom-up processing, and repetition speed. They conclude that the optimal number of repetitions depends on the specific learning targets when using shadowing. They also note the possibility that a stage of automating bottom-up processing may exist and form a necessary basis for the improvement of listening skills through shadowing. In the second full-length article, **Nate Olson** looks at translanguaging practices and investigates how two pairs of Japanese secondary school team teachers implement Soft CLIL. The study reveals complex dynamics in teacher use of Japanese and English, including excessive reliance by teachers on their L1 for translation and explanations, evidence of 'spoon-feeding', and limited scaffolding, suggesting divergence from core CLIL principles by the participants. Olson concludes by highlighting the need for ongoing teacher collaboration, training, and communication for effective application of Soft CLIL principles in context. In the third article, by **Akiko Fujii** and **Yoshinori Inagaki**, readers can gain insight into the use of standardized speaking assessment as a motivational strategy for EFL learners. Their study is informed by ecological classroom practice and assessment, and motivation theory aligned with the L2 Motivational Self System framework. The authors conclude that speaking assessment, assessment feedback, and goal setting might influence the use of motivational strategies, especially when implemented in conjunction with opportunities for feedback and reflection.

Research Forum

The *Research Forum* article by **Imogen Custance** critically evaluates the role of statistical significance testing in quantitative research. Specifically, the author shines light on issues with significance testing and an overemphasis on *p*-values in SLA research. Custance encourages researchers to make efforts to better understand the numbers that underlie tests of statistical significance and to report the accompanying descriptive statistics and effect sizes. The author also makes a case for considering the statistical power necessary for conducting specific tests, and the reporting of p-values that are not statistically significant.

Reviews

In this issue, we are pleased to offer reviews of books about curriculum design; the link between learner corpus research and second language acquisition; and language support for immigrants in Japan from a multicultural community-building perspective.

First, **Kenneth Charles Lambo** reviews *What English Language Teachers Need to Know Volume III: Curriculum, 2nd edition*. According to Lambo, what sets this 2nd edition apart from the earlier one is that "the authors have expanded their chapters to address the pressing technological and multilingual challenges facing learners and teachers in today's English Language Teaching (ELT) education." Although he points out a few shortcomings of the book in terms of providing novice teachers with explicit background information, overall, he regards the book as essential reading for "educators who aspire to make a significant impact in their field, encouraging them to critically reflect on and enhance the delivery and implementation of their language instruction." The second review, written by **Trevor Sitler** is of the book *Learner Corpus Research Meets Second Language Acquisition*, the goal of which is to "maintain that LCR can greatly benefit SLA ... primarily through studies on L1 transfer and proficiency." According to Sitler, one of the benefits that educators can hope to take away from this book is that thanks to advances in technology, educators with only fundamental computer skills can try their hand at corpus research. In the end, Sitler stresses that "this book provides a great introduction to SLA researchers who are looking for a fresh perspective on learner language." In the third review, **Karen Masatsugu** describes *Language Support for Immigrants in Japan: Perspectives from Multicultural Community Building* as a "timely book … a collection of papers about Japanese language education for immigrants in Japan, written by Japanese

academics, teachers, and volunteers for an international English-speaking audience." Although noting a few infelicities in English, she applauds the book for "provid[ing] a rich description of attempts to provide language support to immigrants in Japan through community-based language classes and night schools, the challenges faced and still to be overcome as Japan increasingly depends on immigrant labor, and provides a convincing argument for multicultural community building." As always, we hope that readers will find these reviews helpful.

From the Editors

We, the *JALT Journal* editorial team, extend our sincere gratitude to **William Simpson** who, although having served as Assistant Editor for a short duration, has shared valuable insight and given much-needed support throughout the production of the current issue. Best of luck in your future academic, professional, and personal endeavours, Will. Taking his place is **Paul Leeming**, whose extensive experience as an applied linguist will undoubtedly help ensure that *JALT Journal* evolves as JALT's flagship academic journal, and as one of Japan's most reputable, bilingual contexts for quality research in applied linguistics. Thank you, Paul, for joining our team. Joe and I very much look forward to working with you.

The current issue marks my first as Editor. I owe my deepest gratitude to all my predecessors, notably **Gregory Paul Glasgow** and **Dennis Koyama**, who made my job over the past few years as Assistant and Associate Editor so much easier and rewarding. I am also very much indebted to **Joe Geluso**, our current Associate Editor, whose light-speed and entirely unproblematic adaptation to the job shows what it truly means to be a professional. I only hope I can live up to the standards established by those individuals. With the spotlight on those who have made, and will continue to make, *JALT Journal* the institution that it is, I also feel compelled to highlight the fact that *JALT Journal* editors, as with all JALT officers, organisers and so forth, are volunteers who devote a considerable portion of their work and personal time to creating opportunities for others to present their ideas and research, thus encouraging development in academic research and professional practice. They do so not for personal or professional gain, nor pride, but rather from hope that through collective effort and willingness to learn, we can all face the challenges of our times. Development in our field is also contingent on burgeoning and experienced scholars of all epistemological persuasions prioritising the transformation—rather than the reproduction—of existing scholarly discourse, to take risks, to

embrace the fallibility of academic discourse not as a shortcoming but as a necessary condition for scientific and intellectual development to take place, and to consider the sophistication of existing knowledge in our field as sufficient reward. We, the editors, are honoured to be both spectators and participants in these developments.

JALT Journal remains committed to publishing high-quality research in applied linguistics, especially conceptual and empirically-grounded studies relevant to language education in the Japanese context. We invite readers to read our "Aims and Scope" section at the end of this issue, and consider submitting their research for publication in *JALT Journal*. We also invite our readership to consider submitting special issue proposals. Specific details on the submission process for special issue proposals are available on the JJ website and are printed at the end of the current issue. Considering that academic journals often serve as contexts for academic debate, thus potentially contributing to conceptual and methodological developments, we would also like to bring attention to our journal's *Point-to-Point* section, which offers another opportunity for prospective authors to engage in scholarly debate by commenting on an article previously published in *JALT Journal*. These 1000-word papers are an important part of the journal's contribution to the free exchange of scholarly ideas in our field. The original authors are also invited to follow up with a response to the discussion of their work.

— Jeremie Bouchard, Editor
— Joe Geluso, Associate Editor
— Paul Leeming, Assistant Editor

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

JALT Journal, Vol. 47, No. 1, May 2025

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

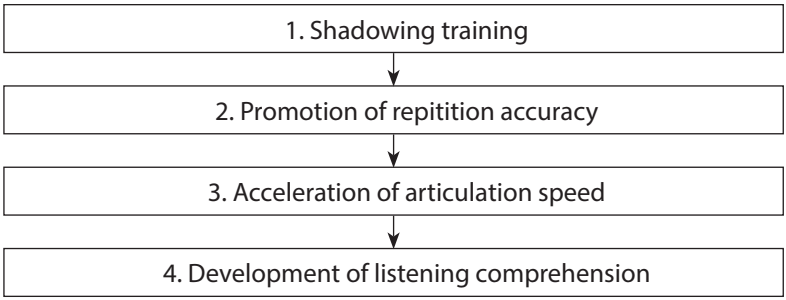
This study focuses on the effects of shadowing on the improvement of L2 listening comprehension (Hamada, 2016a; Tamai, 2005), and the memorization of MWEs (Hashizaki, 2021, 2024c; Miyake, 2009; Xing & Hashizaki, 2021, 2024). Additionally, L2 bottom-up processing (Hamada, 2016a), repetition speed (Tamai,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

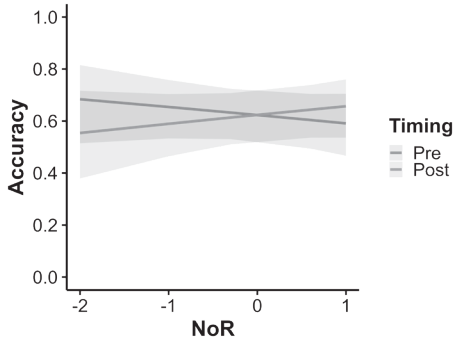
Results

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.

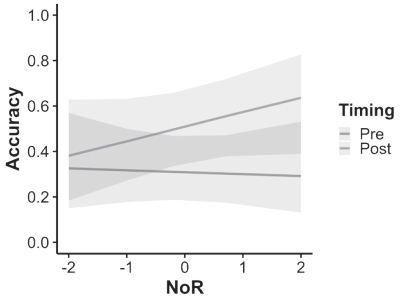
Table 4
GLMM Results for the Listening Test

	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

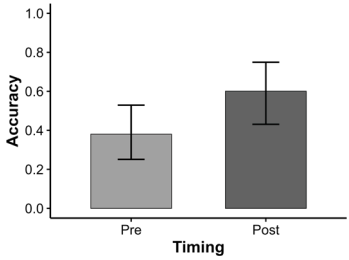
GLMM Results for the Sorting Test

					Random effect	
	Fixed 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

					Random effect	
	Fixed 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	$p = .001$	-	-

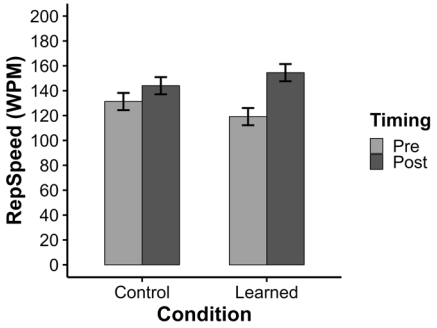
Note. Number of observations = 1,500; $n = 30$. Model formula: Accuracy ~ Timing + (1 | Participant) + (1 + Timing | Item). Timing refers to the time at which the participants took the test. Marginal $R^2 = 0.033$, conditional $R^2 = 0.462$. Timing was simple-coded (Pre = −0.5; Post = 0.5).

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

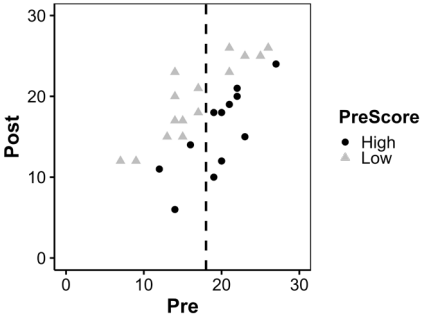
Regarding the listening test, this study's results contrast with those of Hamada (2016a) and Tamai (2005), who found that approximately four to five shadowing repetitions could improve listening comprehension for university students with low- and intermediate-proficiency levels. This can be potentially 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



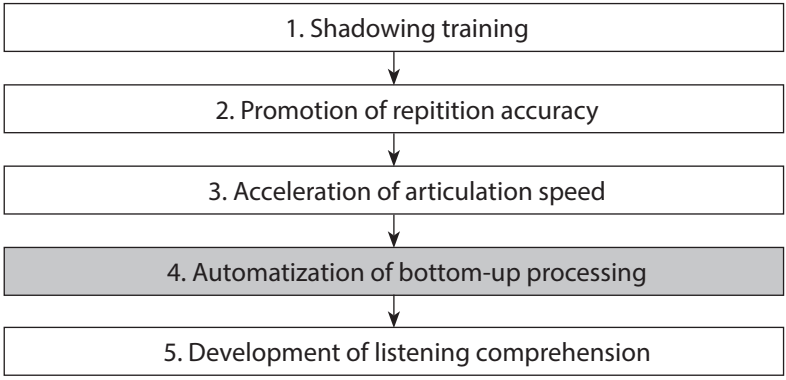
Note. The number of dots totaled 27 (n = 30) because there were three data points that had the same pre- and post-test scores.

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

Scaffolding or Spoon-Feeding? A Case Study of Translanguaging Re-Invention in Team-Taught Soft CLIL Classrooms

Nate Olson
Toyo Gakuen University

This case study examines how 2 pairs of Japanese secondary school team teachers implemented Soft CLIL, with a particular focus on their use of translanguaging. Classroom observations revealed a complex interplay between Japanese (L1) and English (L2), where strategic language integration was evident but often overshadowed by excessive reliance on L1 for translation and explanations. This overuse indicated a re-invention of CLIL that diverged from its theoretical foundations. Teachers justified this adaptation by emphasizing student enjoyment, emotional support, and accommodation of lower proficiency levels. However, misinterpretations of CLIL's principles resulted in frequent "spoon-feeding" rather than effective scaffolding. The findings highlight the need for continuous teacher collaboration, professional development, and clearer communication to ensure fidelity to Soft CLIL's language-learning framework. Without these supports, the risk of re-invention increases, potentially undermining CLIL's intended outcomes. Nonetheless, this study underscores that CLIL remains a promising approach when adapted with awareness of classroom realities.

本事例研究では、日本の中等学校におけるチームティーチングの2組の教師が、ソフトCLILをどのように実施したかを調査し、特にトランスランゲージングの使用に焦点を当てた。授業観察の結果、日本語（L1）と英語（L2）の複雑な相互作用が明らかになり、戦略的な言語統合の試みは見られたものの、L1への過度な依存が翻訳や説明の手段として頻繁に用いられていた。この傾向は、CLILの理論的枠組みから逸脱した「再発明」となっていた。教師たちはこの適応を、生

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

JALT Journal, Vol. 47, No. 1, May 2025

徒の楽しさの確保、感情的サポートの提供、習熟度の低い生徒への配慮として正当化していたが、CLILの原則の誤解によって「足場作り」ではなく「スプーンフィーディング」（過度な手助け）が頻繁に行われていたことが分かった。本研究の結果は、ソフトCLILの言語学習モデルを維持するために、教師同士の継続的な協力、専門的な研修、および明確なコミュニケーションが不可欠であることを

Keywords: secondary school; soft CLIL; team teaching; translanguaging

The growing popularity of Content and Language Integrated Learning (CLIL) has sparked discussions around effective implementation strategies, particularly regarding teacher preparedness (Ikeda et al., 2021; Lo, 2020). Soft CLIL, a flexible approach to CLIL, emphasizes the creation of a language-rich learning environment where students can leverage their existing linguistic resources, including their native language (Olson, 2021). A common feature of Soft CLIL is translanguaging, defined as the dynamic use of two languages "to make meaning, shape experiences, understandings, and knowledge" (Baker & Wright, 2017, p. 280). Using this approach in the Japanese context, students are encouraged to use both Japanese and the target language (typically English) to maximize their learning potential (Ikeda, 2021).

This article addresses the need for research into "the actual in-class effects" of translanguaging (Turnbull, 2018, p. 121) by following two pairs of Japanese secondary school team teachers as they embark on implementing Soft CLIL for the first time. A particular focus is placed on how the team teachers balance the use of L1 (Japanese) and L2 (English) in the classroom and whether "re-invention" (Rogers, 2003) of translanguaging practices occurs during implementation. By investigating these factors, the study aims to contribute insights into the operational realities faced by teachers as they navigate the implementation of translanguaging and Soft CLIL more broadly.

Literature Review

Team Teaching and Soft CLIL

Team teaching in secondary school English language classrooms across Japan has a history extending over three decades. Brumby and Wada (1990, p. 6) describe team teaching in the Japanese context as "a concerted endeavor" where the Japanese teacher of English (JTE) and the assistant language teacher (ALT) collaborate to create a dynamic and communicative learning space. Despite JTEs and ALTs working together for more than 30 years, however, there are still several challenges to effective team

teaching. Some researchers point out a lack of well-established practical and theoretical foundations to support JTEs to adapt and cope with new educational challenges, to prepare ALTs, often new university graduates, to co-teach English in a foreign social and educational environment, and to provide teachers with clear guidelines for team-teaching collaborations (Borg, 2020; Crooks, 2001).

More recently, CLIL has gained traction in Japan as an innovative approach to integrating content and language education. As Coyle et al. (2010) highlight, "what separates CLIL from some established approaches such as content-based language learning, or forms of bilingual education, is the planned pedagogic integration of contextualized content, cognition, communication and culture into teaching and learning practice" (p. 6). CLIL further sets itself apart through its flexibility across a learning continuum. At one end lies "Hard" CLIL, closely adhering to its European origins with academic subjects such as history or science taught predominantly in English by non-native content teachers, with minimal language support. In contrast, "Soft" CLIL offers a more flexible approach, typically led by language teachers (native or non-native speakers) with a stronger emphasis on language learning alongside content acquisition (Ikeda, 2021).

In their survey of Hard and Soft CLIL programs worldwide, Banegas and Hemmi (2021) identify a common emphasis on learner-centered methodology, enhancing critical thinking and Higher Order Thinking Skills (HOTS), and applying translanguaging to improve content comprehension. Ikeda (2019) observes that while Hard CLIL is uncommon in Japanese secondary and tertiary education, Soft CLIL has emerged as the "de facto norm in Japan" (p. 29). Although some scholars caution against simply transplanting European solutions onto Japanese contexts (e.g., Morton, 2019), CLIL advocates in Japan regard it as a transformative educational strategy. They view CLIL as uniquely positioned to equip future generations for new challenges by fostering competencies, pluriliteracies, and enhancing learning experiences through translanguaging (Ikeda, 2021; Tsuchiya, 2019).

Translanguaging in Practice

The concept of translanguaging—the flexible, strategic use of students' full linguistic repertoires in the classroom—has emerged as an innovative approach in language education. Translanguaging has been embraced as a critical response to monolingual English-only policies, challenging the notion that languages should be kept separate in the classroom (García & Wei, 2014). It acknowledges how bi/multilingual learners naturally

integrate their languages and can serve various functions, from enhancing content understanding to facilitating classroom communication (Canagarajah, 2013; Ikeda, 2021). Cenoz and Gorter (2021) distinguish between two types: *pedagogical translanguaging* as a planned instructional strategy integrating students' languages for specific learning aims, and *spontaneous translanguaging* as the natural blending that occurs during interactions.

Although incorporating both types is believed to create a more inclusive environment that leverages learners' full linguistic resources, research suggests teachers are more likely to engage in ad hoc translanguaging without a clear pedagogical plan or awareness of its potential benefits (Wang, 2016). When translanguaging becomes mere spontaneous translation, concerns arise about its consistency and effectiveness for language learning outcomes. As Ikeda (2021) argues, translanguaging "does not mean teachers and learners can resort to their L1 whenever," (p. 88) as this reduces opportunities for target language development. The underlying principle is that translanguaging should be used as a scaffolding strategy to maximize learning. Furthermore, the social justice emphasis of translanguaging in primarily ESL contexts (García & Wei, 2014) may not transfer effectively into the EFL context of Japan (Turnbull, 2021). As speakers of a majority language learning a minority one, Japanese EFL learners may not view themselves as bilingual or embrace the concept of "emergent bilingualism," posing challenges for transferring the ideological aims of translanguaging.

Finally, the lack of teacher training and resources complicates the widespread adoption of strategic, pedagogical translanguaging. Although some studies demonstrate successful implementation with support (e.g., Ikeda, 2019), such cases are limited. Without explicit guidance, in-service secondary school teachers may generally be unaware of translanguaging's aims as an innovative scaffolding approach and struggle to purposefully implement it into their lessons.

Diffusion of Innovations and Re-Invention

In his seminal work *Diffusion of Innovations*, Everett Rogers (2003) proposes a framework for understanding how innovative ideas and technologies disseminate through social systems over time. This theory has found resonance and empirical support across various domains (see, e.g., Peres et al., 2010 for an overview). However, researchers have noted that the process of adopting innovations into real-world practice is often complex (Fullan, 2015; Henrichsen, 1989). The necessity for adjustments to ensure compatibility with existing systems or preferences can lead to

what Rogers termed re-invention: the process whereby users significantly alter or adapt an innovation during its adoption phase (Rogers, 2003).

In the educational domain, re-invention typically entails instructors adapting complex innovations to better align with their own interpretations or misconceptions, thereby making the innovations more applicable and palatable within their specific contexts (Sansom, 2017). These adaptations enable educators to assert autonomy by determining how best to implement innovations in their classrooms. While the resulting re-invention may facilitate quicker adoption rates, it also underscores a fundamental disconnect between theoretical ideals and practical applications, where a theory's original purposes may be overlooked in favor of more immediate, pragmatic solutions.

Re-invention is particularly prevalent with "loosely bundled" innovations like CLIL, which presents broad principles amenable to a wide range of implementations (Ikeda et al., 2021). While CLIL aims to foster enhanced communication and leverage preexisting knowledge by employing techniques such as pedagogical translanguaging, the practical application of these strategies sometimes diverges from their original conceptualizations (Olson, 2023). Within CLIL contexts, translanguaging may veer towards becoming a straightforward means for meaning transference or expeditious translation, straying from its intended pedagogical principles and strategic purposes.

Methodology

This case study investigated how two pairs of Japanese secondary school team teachers implemented Soft CLIL, focusing on their use of translanguaging. The study addressed the following research questions:

RQ 1. How did team teachers balance L1 (Japanese) and L2 (English) use in the classroom?
RQ 2. Did re-invention of translanguaging take place during implementation, and if so, how?

Participants

The participants were part of a larger project examining the collaborative efforts of Japanese secondary school team teachers to adopt and implement Soft CLIL. This study focuses on two educational settings: Take Senior High School (Take SHS) and Ume Junior High School (Ume JHS).

Ethical protocols were followed, including obtaining informed consent and using pseudonyms to protect anonymity.

The participant team at Take SHS consisted of Sato, a male JTE with extensive teaching experience, and Emily, a female ALT from the UK with a decade of experience teaching in Japan. They designed a special elective class to pilot a team-taught CLIL approach, running for 11 weeks with 12 students whose English proficiency ranged from A2 to C1 on the CEFR scale.

The team at Ume JHS was Fujita, a female JTE with over ten years of English teaching experience, and Latoya, an American female ALT in her third year at the school. Their CLIL project was implemented in four lessons over two weeks to a special needs class of four students. Although the students had special needs, these needs were behavioral in nature, and both the researcher and teachers assessed their English proficiency as typical for their grade level, approximately A1 on the CEFR scale.

Collaborative Action Research Approach

Since it was the first time for both teams to implement CLIL, participants were invited to engage in collaborative action research with the researcher throughout the study. Specifically, participants followed a collaborative CLIL teacher development model based on Sasajima (2013) where teachers work together before each class to create lesson plans, materials, and discuss details like teacher roles and scaffolding student needs. After the lesson, teachers reflect on successes, failures, adherence to CLIL principles, roles, student engagement, and other factors. Finally, they revise their approach for subsequent lessons based on this reflection.

To support participants, the researcher provided training resources on CLIL principles and practices before and throughout implementation. These resources covered core CLIL tenets using training videos and templates such as a CLIL Lesson Planning Sheet (Ikeda, 2016), and a Feedback Sheet with checklists for recommended practices (Olson, 2021). However, the onus was on the teachers to review and internalize these resources in their own time, as the researcher's direct involvement was limited to providing the materials and support during meetings.

Data Collection and Analysis

The primary data sources were weekly recorded classroom videos provided by the teachers, as well as recordings from teacher–researcher planning and reflection meetings conducted via Zoom. For Take SHS, the

data comprised 9 hours 13 minutes of recorded classes and 11 hours 47 minutes of teacher meetings. For Ume JHS, it included 3 hours 16 minutes of classes and 5 hours 6 minutes of meetings.

To examine teachers' use of the L1 (Japanese) and L2 (English), quantitative analyses of character/word counts and teacher speaking time were conducted. While all lessons were observed and analyzed qualitatively, a subset of lessons was selected for detailed quantitative analysis based on two key criteria. First, these lessons exemplified critical moments of translanguaging and team teaching that closely aligned with the research focus on how re-invention of CLIL practices occurs. Second, these lessons provided consistently clear audio quality necessary for reliable quantitative analysis, as some recordings had technical limitations that made precise measurement difficult.

As teacher-student interactions during pair/group work were often inaudible, analyses were limited to whole-class, teacher-fronted instruction segments. Following Tsuchiya (2019), an initial quantitative analysis explored overall discourse patterns by transcribing classroom interactions verbatim, excluding fillers. The transcripts were then verified by a Japanese native speaker, timestamped using Transana (Fassnacht & Woods, 2019), and coded using Taguette (Rampkin et al., 2021). Finally, relevant excerpts illustrating teachers' language use and translanguaging practices were selected.

Findings

Take SHS

Sato first learned about CLIL through a former JTE colleague and their efforts to use CLIL at the school. He believed CLIL to be a more authentic method of language education and thought it would help motivate students to learn English. Emily initially learned about CLIL through Sato and did not have any strong feelings about CLIL as an innovation.

Sato and Emily team-taught a total of nine 50-minute CLIL lessons on the theme of Cultural Awareness. After an initial trial lesson on the Philippines, they conducted four lessons on Emily's home country of the UK. For these lessons, Emily prepared the materials and led the content instruction in class. Sato, on the other hand, provided feedback on the materials, added translations, and kept the students on task during the lessons.

Excerpt 1 illustrates how the teachers introduced a worksheet on British stoicism (see Appendix A). In turn 1, Emily attempts to activate

the students' prior knowledge of the concept in English but receives no response, so she calls on a higher-level student instead. However, still not getting a positive response in turn 3, she explains the meaning as written on the worksheet. Sato then signals in turn 4 that he will translate Emily's explanation into Japanese. After translating, he directs the students' attention to the Japanese translation provided at the bottom of the worksheet. Sato tells the students who feel confident in their English to hide the translation, although none of the students in the classroom footage are seen folding over their worksheet or attempting to do so. He then gestures toward Emily, signaling her to continue. In turns 5 and 6, the same pattern is repeated as Emily gives an example of stoicism relating to joy, and Sato again translates it before gesturing for her to proceed.

Excerpt 1

01 Emily: So, today, we are going to talk about stoicism. Stoicism. Does anybody know the meaning? Yes, [S1], you know the meaning? Have you heard of this word before?

02 S1: Never.

03 Emily: Alright, so, stoicism is "not expressing extreme feelings," so it's basically enduring something patiently, putting up with something patiently. So, for example, if you feel pain, or if you feel joy, or if you feel anger, some kind of extreme emotion, you do not show this, okay? This is stoicism.

04 Sato: *Mōikkai īmasu yo. Minasan no nichijō seikatsu nani demo sō desu kedo, sugoi ureshikattari toka, nanka chotto tsurai koto ga atta toki, kanashī koto ga atta toki, sore wa kanjō o dashitai kibun ni narimasu. Dakedo, kono* stoicism *to iu no wa, sō iu kanjō o dasanai. Mā, yoi imi demo warui imi demo, sōiu shugi no, sōiu gaman. Chinami ni, shita ni Nihongo ga arimasu. Moshi hitsuyō na baai, kochira o mite kudasai. Eigo ni jishin aru hito wa kore o kakushitoite ne. Hai, jā, onegaishimasu.* [I'll say it one more time. When you are happy, or when you are in pain, or when you are sad, you feel the need to express your feelings, you know. However, "stoicism" means not showing such emotions. Well, in a good sense and in a bad sense, it is that kind of principle, that kind of patience. By the way, there is Japanese on the bottom [of the sheet]. If you need it, you can look at it. If you are confident in your English, keep this hidden. Yes, well, please [go ahead, Emily].]

05 Emily: So, let's see, about 6 years ago, I went back to the UK because it was my sister's wedding and my older sister got married. So of course I was very happy, and she was standing at the alter and she looks very beautiful, and I started crying. I couldn't stop crying. And my mother was sitting next to me, and she threw a tissue at me. And she said, "stop crying!" So, my mother is very stoic. She didn't want me to show any kind of emotion. Even on my sister's wedding day.

06 Sato: *Yaa, watashi mo kore o kiita toki ni sugoi bikkuri shitan da kedo, roku-nen gurai mae ni, [Emily] no onēsan no kekkonshiki to iu koto de, Igirisu ni modorimashita. Sorede, kekkonshiki-ba no toki ni, saidan no tokoro de, onēsan ga hijō ni kikazatte subarashī sugata de, kō shikijō ni imashita. Sore o mite kandō shite, mō naite shimatta wake desu, atarimae desu ne, soshite tonari ni ita okāsan ga, "nande kono toki naku no?" To tisshupēpā o ban to nagete, "kore de naku no yamenasai!" to iwaretan datte. Chotto bikkuri shita nda ne. Arigatōgozaimasu.* [I was very surprised when I heard this, but about 6 years ago, [Emily] returned to England for her sister's wedding. At the wedding, her sister was there at the altar, looking beautiful in her dress. Seeing this, [Emily] was so moved that she started to cry, and her mother, who was standing next to her, said, "Why are you crying at this moment? and she threw a tissue paper at her and said, "Stop crying over this!" I was a little surprised. Thank you.]

Note. S1 is an identified student

The pattern of Emily providing instructions in English while Sato offers Japanese translations is repeated in Excerpt 2 for the language instruction. In turn 1, Emily reads the examples of language usage directly from the worksheet. Turn 2 shows Sato adding that the term should be familiar since it is also used in Japanese, providing a relevant example for the students, and then gesturing for Emily to proceed. Finally, in turn 3, Emily introduces a question and reads the explanatory passage from the worksheet to answer it. Although not included in the excerpt, Sato subsequently reads aloud the Japanese translation on the worksheet.

Excerpt 2

01 Emily: Alright, so, let's look at the ways we can use this word. So, usage. "He practices stoicism."Okay? "He practices stoicism." The next one: "He is a stoic." And then the last one: "He is stoic."

02 Sato: *Hai, kono tsukaikata no tokoro ne, sono stoicism dato nanka yoku wakaranai tango datta na to omou hito mo, jitsuwa, minasan kiita toko aru to omoimasu. Nihongo ni natte iru ne. Sutoikku. "Kare wa sutoikku da ne." Tatoeba, kyonen sotsugyō shita hito de purogorufā mezashiteita futari mo itan desu yo. Karera wa mō tonikaku sutoikku na seikatsu shite ne, tabemono mo yappari sonna zeitaku mo dekinai desho? Tōzen, jankufūdo, poteto chippusu toka sonna tabenaide, sutoikku na seikatsu o shite… tsumari, nanika o gaman suru, sorede mokuhyō o motte ganbaru, jukensei mo onajida ne, sono hitotachi mo sutoikku ni benkyō shitaita. Hai.* [Yes, this "Usage" part. I think that some of you have heard of "stoicism," even if you think that it is a word that you don't know. It's Japanese, isn't it? *Sutoikku*. He's stoic. For example, there were two students who graduated last year who were aiming to become professional golfers. They lived a stoic life and could not afford to eat extravagantly, could they? They couldn't eat junk food, potato chips, and so on... In other words, they had to endure something, and they worked hard to achieve their goals. Yes.]

03 Emily: Okay, so, if we look below. I have this question: "Why are British stoic?" Okay, so, "Why are British people stoic?"

A vocabulary review activity on the UK content further reveals the teachers' language use in empirical terms, as shown in Table 1. The 10-minute 50-second review had Emily leading the class by reading English fill-in-the-blank sentences and asking students to recall vocabulary words from the previous lesson. Sato then supported her by translating each sentence into Japanese and providing hints about the English vocabulary words. Afterwards, the teachers prompted the students to attempt spelling the words in English.

Table 1
Word Count and Speaking Time Length of Teachers for UK Lesson Vocabulary Review

	Word count		Speaking time		
	English	Japanese	English	Japanese	Total
Sato	31	653	00:42	02:19	03:01
Emily	440	0	03:21	00:00	03:21
Total	471	653	04:03	02:19	06:22

Note. Japanese is shown in characters; Time is shown in minutes and seconds (MM:SS).

During the 10-minute 50-second review, Sato spoke for 42 seconds (31 words) in English and 2 minutes 19 seconds (653 characters) in Japanese, while Emily spoke for 3 minutes 21 seconds (440 words) exclusively in English. Notably, Sato's English usage was limited to repeating the vocabulary words or enunciating them for spelling practice (e.g., "Independence. In-de-pen-dence. Independence."). Emily used English for procedures ("Alright, let's go to the next one."), praise ("Yay, good, good!"), and providing hints ("Very close. Just the end part you need to change."). Conversely, Sato used Japanese for procedures ("*Tsugi no pēji mekutte kudasai*. [Please turn to the next page.]"), praise ("*Oō, subarashi!* [Oh, wonderful!]"), offering hints ("'Expect' *tte ne, kitai suru to iu*. [means expect.]"), and encouragement ("*Machigattemo zenzen ī kara*. [It's totally fine to make a mistake.]"). This division of roles, with Emily leading instruction in English and Sato providing support in Japanese, was a regular pattern in their collaborative lessons.

After Emily had prepared the materials and led the instruction for all four UK lessons, Sato decided to give her a break and take on more responsibilities for the next lesson on Taiwan. He created a worksheet as well as two handouts from online English articles about Taiwan, each including Japanese translations similar to the previous lessons.

During the Taiwan lesson, Sato showed the students a 6-minute Japanese video about Yoichi Hatta, a Japanese engineer who helped build infrastructure in Japanese-occupied Taiwan. His intention was for students to deeply consider Japan-Taiwan relations using the phrase "if possible." However, the classroom footage reveals little effort to enforce or even encourage the use of English. For example, when distributing the first article, Sato

announced to the students: "*Narubeku Nihongo minaide ne. Mā, mitemo ī kedo*. [As much as possible, try not to look at the Japanese. Well, it's okay if you look.]" Minutes later, he admitted: "*Hontō wa Eigo de yaritain desu kedo, nakanaka kore ga Eigo da to muzukashī*. [I really want you to do this in English, but it's quite difficult when it's in English.]"

A quantitative analysis of the lesson, as shown in Table 2, reveals that while Emily used English during the 27 minutes of solo and pair work, Sato conducted the remaining 17 minutes of teacher-fronted, whole-class instruction almost entirely in Japanese. In fact, Sato only spoke English for 1 second, uttering the word "surprise" to indicate where Emily should start reading the article. Excluding her reading aloud to the class, Emily's English usage was limited to 11 seconds (18 words) when preparing to read the article ("Okay, so...") and briefly at the end regarding the homework ("Did anyone do the homework from last time? Oh, you did? Thank you. Perfect, perfect. Yay!").

Table 2

Word Count and Speaking Time Length of Teachers for Taiwan Lesson

	Word count		Speaking time		
	English	Japanese	English	Japanese	Total
Sato	1	3854	00:01	16:35	16:36
Emily	18	0	00:11	00:00	00:11
Total	19	3854	00:12	16:35	16:47

Note. Japanese is shown in characters; Time is shown in minutes and seconds (MM:SS).

During the reflection meeting, Emily expressed frustration at being excluded from the preparations, stating: "I didn't receive a lesson plan or anything, so I didn't fill out the other sheet, the feedback one, because there was no lesson plan... I received the article, but I didn't know what we were doing with it." She further elaborated:

> I think I should have had a more active role because I felt like I was just standing and listening most of the time. I know this lesson was different from usual, but I would have liked to know what we were doing next. With the student worksheet, I didn't see that in advance, so I had to stand there, read it, and figure out what it was saying.

Despite her intermediate Japanese proficiency, Emily admitted having difficulty understanding and "zoning out" during the video and Sato's Japanese lecture on the content. Although Sato had assumed a leader role for this lesson, he acknowledged his mistake in a pre-meeting message: "*Konkai no jyugyō wa, Taiwan to Nihon no kankei ni tsuite mazu Nihongo de kangaesaseru koto wo mokuteki to shimashita node*, [Emily] *Sensei no yakuwari ga sukunakunatte shimaimashita*. [For this class, my objective was to have the students think about the Taiwan-Japan relationship first in Japanese, so Emily's role was regrettably diminished.]" Sato's intention to take more responsibility and not overburden Emily during planning seemed to backfire. However, they learned from this experience and went on to teach four lessons on the US in a more collaborative manner that aligned with their approach to the UK lessons.

During the final teacher interviews, Sato explained that he adopted CLIL because he wanted to teach more challenging content and use translanguaging as a scaffolding method without relying solely on English. This stemmed primarily from his concerns about the students' varying English proficiency levels and his prioritization of ensuring everyone could understand and enjoy the content. When asked about strategic translanguaging, he admitted: "*Yahari riron wa subarashī ga jissen ni kanshite wa, sōtō no jyunbi to doryoku ga hitsuyō ni naru*. [The theory is excellent, but putting it into practice requires a lot of preparation and effort.]" Emily also acknowledged, "With CLIL, there are so many things to consider in order to have a 'successful' lesson." Sato seemed to agree, stating at one point: "*Mesoddo ga shikkari shiteitemo, sore wo namami no kōkōsei ni oshitsuke ni naranai yō ni kufū shimashita*. [Even with solid methods, I had to find ways to ensure that CLIL would not feel imposed upon the high school students.]" Considering Take SHS's relaxed attendance policy, his stated goal was "*narubeku doroppuauto shinai yō ni, tanoshinde morau yō ni shimashita*. [To have the students enjoy the team-taught CLIL class as much as possible without dropping out.]" Finally, when asked about the future use of CLIL at Take SHS, Sato said "*seito no kyōmi to nōryoku ni ōjite tsukaitsudzukeru tsumori desu*. [We plan to continue using it based on the students' interests and abilities.]"

Ume JHS

Similar to Sato, Latoya viewed CLIL as an opportunity to teach content without solely relying on English. By her own assessment, Latoya was already quite knowledgeable about CLIL practices, having implemented

CLIL-like projects with special needs students in the past. For the present study, Latoya convinced Fujita to team-teach four 50-minute CLIL lessons on the theme of Cultural Awareness. Fujita agreed to let Latoya plan and lead the content instruction while providing feedback and assisting as the classroom manager.

For the first lesson, Latoya wrote the target language on the board ("Where do you want to go?"; "What do you like?"; "What do you want to do?"), and the teachers modeled answering the questions, as shown in Excerpt 3. Turn 2 not only reveals Fujita's Kansai dialect but also illustrates how she often mixed Japanese and English in her speech, seemingly to maintain the students' attention, as evident in turns 15 and 20. After turn 14, as Latoya began loading a PowerPoint presentation example with her back to the class, Fujita kept her eyes on the students and directed their attention in turn 18. This dynamic, with Latoya leading the procedures and Fujita fulfilling her role as classroom manager, was representative of their team-teaching roles for the remainder of the project.

Excerpt 3

01	Latoya:	Ms. [Fujita], where do you want to go?
02	Fujita:	*Yutta kamoshirehen kedo* [I might have already said it, but] I want to go to Finland.
03	Latoya:	Woah, you want to go to Finland? Why?
04	S1:	Why go?
05	Fujita:	Because, do you know Moomin? I like Moomin very much.
06	S1:	I like Moomin. Mother, mother, I love you. (laughs)
07	Fujita:	(laughs) Moomin is from Finland. So, I want to visit Finland.
08	S1:	*Shusshinchi? Mūmin no shusshinchi wa Finrando?* [Hometown? Moomin's hometown is Finland?]
09	Fujita:	I want to see *ōrora*. [the aurora (borealis).]
10	S1:	*Ōrora wo mitai.* [You want to see the aurora (borealis).]
11	Latoya:	*Ōrora raitsu?* [The aurora lights?]
12	S1:	♪*Kirakira kirakira…* [Twinkle, twinkle…] ♪
13	Fujita:	Very good, very beautiful. So, that's why I want to go to Finland.
14	Latoya:	So, you like Moomin. And you like pretty lights. So you want to go, you want to see Moomin museum in Finland? And you want to go see aurora lights in Finland?

15 Fujita: Yes. *O,* [Latoya] *Sensei ni kiite miru ka?* [Oh, should we ask [Latoya]?] [Latoya], where do you want to go?
16 Latoya: I want to go to…
17 S1: I want to go to…
18 Fujita: Look at the screen.
19 Latoya: Vietnam. Do you know Vietnam?
20 Fujita: *Wakaru?* [Do you know?] Do you know? *Nihongo de wa "Betonamu" ne* ["Vietnam" in Japanese, right?]

Note. S1 is an enthusiastic female student.

The second lesson was the final one with sustained teacher-fronted, whole-class instruction. Excluding the students' solo work time researching countries they wanted to visit, the teachers led classroom activities for 19 minutes and 15 seconds. Table 3 summarizes the teachers' word count and speaking time during this period. Compared to a similar total speaking duration for an activity at Take SHS (cf. Table 1), the L1 and L2 usage appeared more balanced between Latoya and Fujita. While Fujita spoke English for 42 seconds, the same as Sato, she used more than double the number of words, indicating a faster pace of speech. Notably, unlike Emily, Latoya also spoke Japanese and had significantly more speaking time than Fujita, suggesting she took on more of a leadership role.

Table 3

Word Count and Speaking Time Length of Teachers for Lesson 2 (Country Comparison)

	Word count		Speaking time		
	English	Japanese	English	Japanese	Total
Fujita	77	529	00:42	01:51	02:33
Latoya	367	353	02:45	01:17	04:02
Total	444	882	03:27	03:08	06:35

Note. Time is shown in minutes and seconds (MM:SS); Japanese is shown in characters.

In line with CLIL principles, Fujita provided instruction focused on both topic knowledge (e.g., "[Latoya] *Sensei wa doko ni ikitain? Betonamu ya na. Ī na. Metcha shashin kirei ya na.* [Where does Latoya want to go? Vietnam,

right? That's nice. The pictures look really pretty, don't they?]") and metalanguage (e.g., "What day? *Nan yōbi?* [What day of the week?]"). Latoya also offered metalanguage instruction (e.g., "Before we talk [sic] about 'I want to go.' 'I want to go to *nani nani* [something something]'"). However, unlike Emily, Latoya's topic knowledge instruction was primarily in Japanese. For example, when providing procedures for comparing countries, she said:

> *Hitobito wa onajiku nai ne. Sō desu kara, minasan no kuni wa erabeta no kuni ni Nihon to sono kuni wa nani ga chigaimasu ka? Sono kami ni kaita hō ga ī. Nihon wa chigau to, sono kuni wa chigau ne. Hai, minasan, Nihongo de kaite kudasai.*

While her meaning may have been conveyed to the students (that people are different, so they should write in Japanese on their worksheets what distinguishes their chosen country from Japan), it is worth noting that she made some noticeable grammatical errors in Japanese. For instance, "not the same" should be *onaji jyanai* instead of *onajiku nai*, and "chosen country" should be *eranda kuni* rather than *erabeta no kuni*.

After providing the procedures, the teachers handed out the students' worksheet (see Appendix B). It is noteworthy that the worksheet was written entirely in Japanese, with prompts such as "What are the differences between your country and Japan?", "What are the similarities between your country and Japan?", and "What do you think of your country?". Moreover, in the classroom footage, when one female student attempted to write a response in English, Latoya can be seen erasing her sheet while saying: "*Eigo wa muzukashisugiru.* [English is too difficult.]" During the reflection meeting, Latoya elaborated on this incident, stating that "Japanese is better [for her] because [she] can explain more and express more things."

The students then created PowerPoint slides and presented on their chosen countries (Italy, Korea, the US, and China) using a mixture of English and Japanese. Reflecting on the project, Fujita commented that the students' presentations were wonderful, elaborating: "*Hitomae de hanasu koto ga nigate na seito mo ita ga, yarikiru koto ga dekite jishin ni natta to omou*. [Some students are not very good at public speaking, but I think they gained confidence by overcoming this challenge]." Latoya, providing her overall impression, reflected:

> I think this project ended up going very well. It has helped reinforce certain grammar points and aided with public speaking. It was really interesting helping the students

> discover what they liked about different countries and the experiences they can try.

When asked about the difficulties in implementing the project, Latoya commented that "some of the challenges were explaining to Fujita about using CLIL and how Japanese can be and sometimes should be used in the classroom." Reflecting on this, Fujita stated, "*Amari takusan tsukau to yokunai desu kedo, seito ga, 'n?' to natta toki ni, Nihongo de chotto hitsuyō na bubun dake ittekuretari suru koto de kodomotachi ga anshin suru kana to omotteimasu.* [It's not advisable to use Japanese too much, but I think it puts the children at ease if the ALT can say a little bit in Japanese when they seem confused.]" Ultimately, however, Fujita believed that future CLIL projects at the school would be logistically challenging. She preferred the more structured textbook activities, as the open-ended nature of the project was too chaotic for the special needs students.

Discussion

Balancing L1 and L2 in Team-Taught CLIL

The classroom excerpts and quantitative analyses revealed intricate dynamics in how the team teachers balanced their use of Japanese (L1) and English (L2) during CLIL instruction. At Take SHS, a distinct pattern emerged where Emily typically spearheaded content explanations and procedural instructions in English, while Sato provided corresponding Japanese translations and facilitated classroom management strategies like encouragement and praise. This role division aligns with traditional team teaching dynamics identified in prior research (e.g., Brumby & Wada, 1990). However, a pronounced imbalance became evident in Sato's Taiwan lesson, where he conducted nearly the entirety of the teacher-fronted instruction in Japanese, effectively relegating Emily to a marginalized role of merely reading English scripts aloud. This extreme case exemplifies the "human tape recorder" phenomenon reported by disenfranchised ALTs in previous studies (Borg, 2020; McConnell, 2000), marking a lack of not only team teacher collaboration but also individual teacher agency.

At Ume JHS, a more balanced overall utilization of L1 and L2 was observed between Latoya and Fujita during whole-class instructional segments. Yet Latoya's relatively low Japanese proficiency was evidenced through errors in her instructions, which may inherently limit her ability to strategically and smoothly transition between languages as a means to scaffold student learning. This observation suggests that a higher level

of L1 proficiency may be a necessary prerequisite for ALTs to effectively employ scaffolding techniques through the students' first language. Furthermore, Latoya's insistence on using Japanese as the primary medium to explore cultural content in-depth, coupled with her discouragement of a student's attempt at L2 writing, suggests a concerning reversion to positioning Japanese as the default language. Under typical circumstances within the CLIL framework, students may indeed be permitted to utilize their L1 strategically to enhance the Cognition component of the approach; however, this would be implemented with the underlying expectation that the final output and production be conducted in English (see, e.g., Ikeda, 2016).

Overall, these findings underscore how team teachers' respective language proficiencies, particularly in the L1, as well as the presence (or lack thereof) of substantive collaborative planning can significantly impact their ability to strategically integrate both languages in adherence to core CLIL principles. Even with sufficient training and opportunities for collaborative lesson planning, an excessive dependence on the L1 may emerge, potentially compromising opportunities to challenge students and promote growth in the target language.

Reconciling Scaffolding Aims and Re-Inventive L1 Reliance

The data indicates that re-invention of translanguaging practices did indeed occur during the process of CLIL implementation by both teacher pairs examined in this study. At Take SHS, while Sato initially expressed intentions to leverage translanguaging as a strategic scaffolding approach, his pedagogical priorities appeared to shift over time towards primarily ensuring that students could access the lesson content and "enjoy" classroom activities without being overburdened. This realignment of aims resulted in a tendency to resort to extended explanations exclusively in Japanese as well as instances of ad hoc translation from English to Japanese. Although translanguaging proponents argue that spontaneous translanguaging can be used in a pedagogically productive manner (Cenoz & Gorter, 2020), the observations in this case study revealed a clear overreliance on this ad hoc type of language integration practice to the detriment of more planned, strategic implementations of pedagogical translanguaging. Similarly, at Ume JHS, Latoya expressed a belief that utilizing Japanese was a necessity in order to fully explore the nuances of cultural content, despite her persistent struggles to clearly articulate a coherent pedagogical rationale to her team-teaching partner Fujita. This perspective aligns with Wang's

(2016) observation that teachers tend to frequently engage in spontaneous translanguaging practices that lack intentional strategic aims.

Furthermore, the contextual reality that both schools opted to implement CLIL in specialized educational settings—an elective cross-grade class and a class dedicated to students with special needs—rather than within mainstream English language classrooms, may have contributed to re-inventive tendencies that effectively deprioritized strict adherence to CLIL's established principles in favor of more immediate aims like ensuring student enjoyment and avoiding potential frustration.

As Garton and Copland (2021) assert, however, it cannot be the case that "anything goes" (p. 5) with CLIL classroom practices. The findings suggest that the relatively loose definitions surrounding the concept of "translanguaging" appear to have inadvertently enabled these re-inventive practices centered on excessive reliance on the L1, running contrary to CLIL's core principles. Moving forward, advocates and practitioners of the CLIL approach should remain cognizant of the risk that a wholesome rejection of restrictive "English-only" instructional policies does not become re-invented into an equally unproductive ethos of "Japanese-only" within nominally English-focused classrooms, as was exemplified in the extreme case of Sato's Taiwan lesson.

Ultimately, while the existing literature highlights translanguaging's potential as a theoretically sound approach to scaffolding linguistic development, the findings of the present study demonstrate how the realities of actual classroom implementation can lead to re-inventive practices that substantially loosen the boundaries between judicious, strategic linguistic integration and excessive, unnecessary reliance on students' L1. As Ikeda (2021) cautioned, such re-inventive overuse of the L1 risks reducing rich opportunities for productive linguistic development in the very target language that CLIL aims to cultivate.

Conclusion

This case study examined how two pairs of Japanese secondary school team teachers navigated the implementation of Soft CLIL, with a particular focus on their use of translanguaging practices. The findings revealed complex dynamics and imbalances in how the teachers utilized Japanese and English during CLIL lessons. While some attempts at strategic language integration were observed, instances of excessive L1 use for translation and content explanations suggested re-invention (Rogers, 2003) that strayed from CLIL's theoretical foundations. Teachers' rationales for

re-invention included ensuring student enjoyment, providing emotional support for learners, and accommodating lower proficiency levels.

Misinterpretations about CLIL's aims for scaffolding and judicious translanguaging were also evident. Despite CLIL's emphasis on offering support only when necessary through scaffolding, several examples of "spoon-feeding" were documented. The findings underscore the need for ongoing collaboration, training, and clear communication between team teachers to maintain fidelity to Soft CLIL's model of language learning. Without these supportive conditions, the risk of re-invention and ad hoc implementation increases, potentially undermining CLIL's core tenets.

A significant limitation of this study was the teachers' relatively limited training in CLIL principles and translanguaging pedagogy, which likely influenced their implementation practices. However, this limitation itself reveals an important finding about how educational innovations are typically adopted in real-world contexts—often with incomplete understanding that leads to re-invention. This aligns with Rogers' (2003) diffusion of innovations theory, which recognizes that practitioners frequently modify new approaches to fit their specific circumstances and understanding. Future research would benefit from comparing implementation patterns between teachers with varying levels of CLIL training to better understand how professional development impacts fidelity to the approach's core principles.

After reviewing a decade of CLIL implementation across Europe, Georgiou (2021) found that "the CLIL umbrella might be stretching too much" (p. 497). However, she concluded: "It is clear that CLIL, as an innovation, was difficult to implement perfectly at the beginning, but that should not deter us from striving towards improving an approach that has important potential for language learning and education in general." The present study's findings support this perspective—while the challenges of implementation are substantial, they should not overshadow CLIL's transformative potential. Rather, these challenges highlight the importance of developing comprehensive teacher training programs, creating clear implementation guidelines, and fostering sustained communities of practice where teachers can collaboratively work through the complexities of CLIL adoption. Through such systematic support and continued research into actual classroom practices, CLIL's vision for integrated content and language learning can be more effectively realized in Japanese secondary education.

Nate Olson is an Assistant Professor at Toyo Gakuen University. His research interests include CLIL, team teaching, teacher development, and communities of practice.

Appendices

All appendices are available from the online version of this article at https://jalt.org/main/jj.

References

Baker, C., & Wright, W. E. (2017). *Foundations of bilingual education and bilingualism* (6th ed.). Multilingual Matters. https://doi.org/10.1007/978-3-319-02258-1_2

Banegas, D. L., & Hemmi, C. (2021). CLIL: Present and future. In C. Hemmi & D. Banegas (Eds.), *International perspectives on CLIL* (pp. 281–295). Palgrave Macmillan. https://doi.org/10.1007/978-3-030-70095-9_14

Borg, P. (2020). The JET Programme's de facto 'ESID' policy and its consequences: Critical perspectives from the online 'ALT community'. *Gifu Kyoritsu University Ronshu*, *53*(3), 41–59.

Brumby, S., & Wada, M. (1990). *Team teaching*. Longman.

Canagarajah, A. S. (2013). *Translingual practice: Global Englishes and cosmopolitan relations*. Routledge. https://doi.org/10.4324/9780203073889

Cenoz, J., & Gorter, D. (2021). *Pedagogical translanguaging*. Cambridge University Press. https://doi.org/10.1017/9781009029384

Coyle, D., Hood, P., & Marsh, D. (2010). *CLIL: Content and language integrated learning*. Cambridge University Press. https://doi.org/10.1017/9781009024549

Crooks, A. (2001). Professional development and the JET Program: Insights and solutions based on the Sendai City Program. *JALT Journal*, *23*, 31–46. https://doi.org/10.37546/JALTJJ23.1-2

Fassnacht, C., & Woods, D. (2019). *Transana* (Version 3.32) [Mac]. https://www.transana.com/

Fullan, M. (2015). *The new meaning of educational change* (5th ed.). Teachers College Press.

García, O., & Wei, L. (2014). *Translanguaging: Language, education, and bilingualism*. Palgrave Macmillan. https://doi.org/10.1057/9781137385765_2

Garton, S., & Copland, F. (2021). Series editor preface. In C. Hemmi & D. Banegas (Eds.), *International perspectives on CLIL* (pp. v–vii). Palgrave Macmillan.

Georgiou, S. I. (2012). Reviewing the puzzle of CLIL. *ELT Journal*, *6*(4), 495–504. https://doi.org/10.1093/elt/ccs047

Henrichsen, L. E. (1989). *Diffusion of innovations in English language teaching: The ELEC effort in Japan, 1956–1968*. Greenwood.

Ikeda, M. (2016). CLIL活用の新コンセプトと新ツール [New concepts and new tools to activate CLIL]. In M. Ikeda, Y. Watanabe, & S. Izumi (Eds.), *CLIL (Content and language integrated learning): New challenges in foreign language education at Sophia University* (Vol. 3, pp. 1–29). Sophia University Press.

Ikeda, M. (2019). CLIL in comparison with PPP: A revolution in ELT by competency-based language education. In H. Reinders, S. Ryan, & S. Nakamura (Eds.), *Innovation in language teaching and learning: The case of Japan* (pp. 23–45). Palgrave Macmillan. https://doi.org/10.1007/978-3-030-12567-7_3

Ikeda, M. (2021). Pedagogical translanguaging in primary school math CLIL lessons in Japan. In M. Makoto, S. Izumi, Y. Watanabe, R. Pinner, & M. Davis (Eds.), *Soft CLIL and English language teaching: Understanding Japanese policy, practice, and implications* (pp. 86–102). Routledge. https://doi.org/10.4324/9780429032332-5-6

Ikeda, M., Izumi, S., Watanabe, Y., Pinner, R., & Davis, M. (Eds.). (2021). *Soft CLIL and English language teaching: Understanding Japanese policy, practice, and implications*. Routledge. https://doi.org/10.4324/9780429032332

Lo, Y. Y. (2020). *Professional development of CLIL teachers*. Springer. https://doi.org/10.1007/978-981-15-2425-7

McConnell, D. (2000). *Importing diversity: Inside Japan's JET Program*. University of California Press. https://doi.org/10.1525/9780520921931

Morton, T. (2019). Foreword: CLIL as transgressive policy and practice. In K. Tsuchiya & M. D. Pérez Murillo (Eds.), *Content and language integrated learning in Spanish and Japanese contexts: Policy, practice and pedagogy* (pp. v–xii). Palgrave Macmillan. https://doi.org/10.1007/978-3-030-27443-6

Olson, N. (2021). Tools for team-taught CLIL implementation. *Journal of the Japan CLIL Pedagogical Association*, *3*, 10–30.

Olson, N. (2023). Innovation theory for team-taught soft CLIL: A case study of junior high school adoption and implementation. *Asian Journal of Content and Language Integrated Learning*, *1*, 29–44.

Peres, R., Muller, E., & Mahajan, V. (2010). Innovation diffusion and new product growth models: A critical review and research directions. *International Journal of Research in Marketing*, *27*(2), 91–106. https://doi.org/10.1016/j.ijresmar.2009.12.012

Rampkin, R., Steeves, V., & DeMott, S. (2021). *Taguette* (Version 0.10.1) [Mac]. https://doi.org/10.5281/zenodo.4560784

Rogers, E. M. (2003). *Diffusion of innovations* (5th ed.). Free Press.

Sansom, D. (2017). Reinvention of classroom innovations. *ELT Journal*, *71*(4), 423–432. https://doi.org/10.1093/elt/ccw116

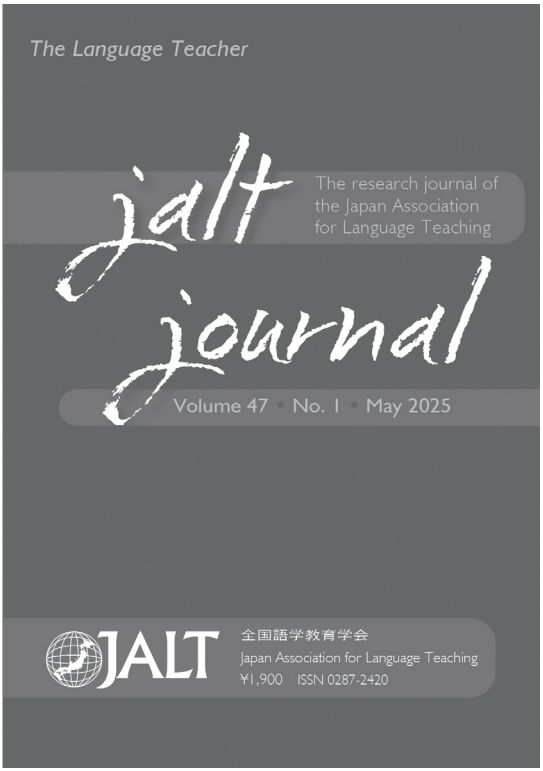
Sasajima, S. (2013). How CLIL can impact on EFL teachers' mindsets about teaching and learning: An exploratory study on teacher cognition. *International CLIL Research Journal*, *2*(1), 55–66.

Tsuchiya, K. (2019). Translanguaging performances in a CLIL classroom at a Japanese university. In K. Tsuchiya & M. D. Pérez Murillo (Eds.), *Content and language integrated learning in Spanish and Japanese contexts: Policy, practice and pedagogy* (pp. 263–284). Palgrave Macmillan. https://doi.org/10.1007/978-3-030-27443-6_11

Turnbull, B. (2018). Is there a potential for a translanguaging approach to English education in Japan? Perspectives of tertiary learners and teachers. *JALT Journal*, *40*(2), 101–134. https://doi.org/10.37546/JALTJJ40.2-3

Turnbull, B. (2021). Am I bilingual? Reporting on the self-reflections of Japanese EFL learners. *International Journal of Bilingualism*, *25*(5), 1–22. https://doi.org/10.1177/13670069211019467

Wang, D. (2016). Translanguaging in Chinese foreign language classrooms: Students and teachers' attitudes and practices. *International Journal of Bilingual Education and Bilingualism*, *22*(2), 138–149. https://doi.org/10.1080/13670050.2016.1231773



JALT Journal Online

https://jalt-publications.org/jj

47 Volumes
87 Issues
Hundreds of articles!
1979~2025

Transforming Motivation Into Motivated Behavior: The Role of a Standardized Speaking Assessment as a Strategy for Understanding the Current L2 Self

Akiko Fujii
International Christian University
Yoshinori Inagaki
Tsuda University

The aim of the current study is to explore the use of a standardized speaking assessment as a motivational strategy for EFL learners within the L2 Motivational Self System (L2MSS) framework. The study integrates an ecological classroom practice, assessment, with motivation research based on the L2MSS and focuses on how learners perceive their current L2 self, which has been pointed out to be an under-researched aspect of designing motivation interventions. The assessment was included in a pedagogical intervention cycle and was intended to provide learners with an understanding of their current L2 self and the discrepancy between their current and future-self, thus enabling learners to map a realistic path to their goals. Trajectories of learners' motivation suggest that the pedagogical intervention did impact learners' self-perceptions of their speaking ability as well as the nature of their actual motivated behavior. Therefore, classroom assessment may be an effective motivational strategy, especially when implemented in conjunction with opportunities for feedback and reflection.

https://doi.org/10.37546/JALTJJ47.1-3

JALT Journal, Vol. 47, No. 1, May 2025

本研究の目的は、L2動機づけ自己システム（L2MSS）の理論的枠組みに基づき、EFL学習者の動機づけを高めるストラテジーとして、スピーキング評価の活用を検討することである。本研究は、生態学的教育実践の一つである「評価」をL2MSSに基づく動機づけ研究に統合し、学習者がL2使用者としての現在の自己像をどう認識しているかに焦点を当てるものであり、学習者の動機づけへの介入に関するこれまでの研究において十分に取り上げられてこなかった点に注目するものである。本研究において、スピーキング評価は一連の教育的介入サイクルに組み入れられ、学習者にL2使用者としての現在の自己像を、そしてまたその自己像と自分が目指す自己像との差を認識させることで、学習者に目標達成までの現実的な道筋を描けるようにすることを狙うものであった。学習者の動機づけに見られた変化は、この教育的介入が学習者のスピーキング能力に関する自己認識だけでなく、実際の学習行動のあり方にも影響したことを示唆している。このことから、教室内における評価は、特にフィードバックと振り返りの機会を伴った場合、効果的な動機づけストラテジーとして機能する可能性があるといえる。

Keywords: current L2 self; L2 motivational self-system; motivation; speaking assessment

In a foreign language context, where learners often do not have sufficient opportunities for practical use of the target language, motivation may be one of the most fundamental factors that determines progress in second language learning. Yet, even when learners are motivated in the sense that they feel a strong desire to improve their second or foreign language (L2) skills, they may still fail to engage in the instructional learning activities or make the necessary efforts to improve their skills. Instructors need to understand not only how to raise learners' motivational levels, but also how to motivate learners to actually engage in effective learning behaviors. This point has been emphasized in studies that have focused on motivational teaching practices and their outcomes in the form of actual motivated learning behavior (Guilloteaux & Dörnyei, 2008; Papi & Abdollahzadeh, 2012; Sato, 2021). Specifically, an increasing number of studies have applied Dörnyei and Kubanyiova's (2014) vision-centered teaching practices to explore the power of future visions in enhancing learners' motivation and motivated learning behavior (Le-Thi et al., 2022; Magid & Chan, 2012; Safdari, 2021; Sato, 2021; Sato & Lara, 2019). However, there is still much that remains unknown about the mechanism of future visions. For example, it has been argued that there is still insufficient understanding of one of the crucial dynamics within the L2 Motivational Self System (L2MSS) model (Henry & Cliffordson, 2017; Thorsen et al., 2020)—the discrepancy between current and future self-guides. The current paper addresses this under-researched dimension by exploring the role of assessment practices as a strategy to bridge this gap between current and future self-guides. It is proposed that assessments in the classroom can help to develop learners' awareness of their current L2 self, which in turn can then provide learners with a concrete base for mapping out a realistic plan to achieve their future vision.

Literature Review

Motivational Strategies in the L2 Classroom

Motivation has been defined as "a general way of referring to the antecedents (i.e., the cause and origins) of action" (Dörnyei, 2001, p.6), including both the reason why an action is taken, as well as the effort and persistence, or intensity associated with the action. Dörnyei (2001) also introduced the concept of "motivational strategies" to refer to "techniques that promote the individual's goal-related behavior" (p. 29) and outlined a framework for understanding motivational teaching practices in the L2 classroom.

In the past two decades, a large number of studies have been conducted to investigate motivational teaching practices in the L2 classroom (see Boo et al., 2015, and Lamb, 2019 for reviews). Dörnyei and Csizér (1998) asked Hungarian teachers of English to rank motivational strategies in the order of their perceived importance and identified a list of 10 macrostrategies, otherwise known as the "Ten Commandments for Motivating Language Learners." They also found that "increasing learners' goal-setting" was one of the most underused strategies by teachers. Since then, a number of other studies such as Cheng and Dörnyei (2007) and Sugita and Takeuchi (2010) have also identified potentially effective motivational techniques. Notably, Guilloteaux and Dörnyei (2008) addressed the need for concrete evidence of the effects of motivational practices on motivated behavior by basing their findings on actual observable behavior. Quite a number of further studies were conducted to investigate the relationship between motivational strategies and learners' motivation (e.g., Le-Thi et al., 2022; Moskovsky et al., 2013; Papi & Abdollahzadeh, 2012). However, one limitation of previous research is the tendency to focus generally on the effect of teacher's instructional practices rather than on specific motivational strategies and specific cognitive processes (Lamb, 2017; Le-Thi et al., 2022; Ushioda, 2016). Therefore, the current study responds to the need to investigate particular motivational strategies in more detail, including their classroom implementation and outcomes, by focusing on one specific motivational strategy, assessment.

Assessment as a Motivational Strategy

Few studies within the field of L2 motivation make explicit reference to assessment as a motivational strategy (Gan et al., 2019). Yet, current formative approaches to assessment highlight the crucial role of testing

as a learning tool (e.g., Chong & Reinders, 2023; Fox, 2014; Gebril, 2023). For instance, learning-oriented assessment (Carless, 2015) focuses on "the potential to develop productive student learning processes" (p.964), and emphasizes the fundamental role of self-evaluation and engagement with feedback. Similarly, descriptions of diagnostic uses of language testing (Jang & Wanger, 2014; Kissling & O'Donnell, 2015) also emphasize the role of feedback on assessments, which can benefit learning by helping learners to recognize "the gap between the learners' current level of performance and a desired level of performance or goal" (Jang & Wanger, 2014, p. 698). For example, it was found that the use of self-assessment of oral performance following the ACTFL Oral Proficiency Guidelines led to greater language awareness and self-efficacy (Kissling & O'Donnell, 2015). The potential role of assessment in helping learners to visualize concrete learning goals is consistent with current trends in pedagogical interventions based on the L2 Motivational Self-System.

The L2 Motivational Self-System (L2MSS) and Motivated Learning Behavior

One of the main frameworks employed by recent studies on L2 motivation is Dörnyei's L2MSS (Csizér, 2019; Dörnyei, 2005, 2009), which consists of three components: (1) the ideal L2 self, which is related to the desire to reduce the discrepancy between the actual self and ideal self, (2) the ought-to L2 self, which is related to learners' views of what they fear or want to avoid becoming, and (3) the L2 learning experience, which relates to the effect of the learners' immediate learning environment such as the teacher, curriculum, and experience of success. The ideal L2 self and ought-to self, also referred to as future self-guides, can serve as motivating forces, especially when the self-guide is accompanied by an elaborate and vivid self-image. A growing number of studies have been conducted to explore the practical implications of the L2MSS framework for the classroom (e.g., Csizér & Kormos, 2009; Hiver & Al-Hoorie, 2020; Lamb, 2012; Yashima et al., 2017; You et al., 2016), including the use of the L2MSS framework to investigate the motivational effects of teaching practices. In an Iranian EFL context, Papi and Abdollahzadeh (2012) found a strong correlation between teachers' motivational practices and students' motivated behavior. However, in investigating the relationship between students' ideal L2 selves and motivated behavior, they found no difference between the high motivation and low motivation groups with regard to learner's ideal L2 selves, concluding that "only having an imaginary picture of one's desired L2 self cannot result in actual

motivated behavior unless conditions are met and decisive steps are taken to facilitate realizing the ideal L2 selves" (p.590).

In fact, this is a point that was addressed by Dörnyei (2009) from the start, and further articulated in Dörnyei and Kubanyiova's (2014) framework of vision-centered teaching practices. The framework outlined six conditions that increase the impact that the ideal and ought-to self may have on learners' motivated behavior: (1) learners possess a future self-image, (2) the vision is elaborate and vivid, (3) the future self-image is perceived as realistic or "plausible," (4) learners have some concrete action plan which specifies the steps needed to achieve their goal, (5) the vision is regularly activated, and (6) the learner has an image of undesirable negative consequences for not attaining the ideal self.

Particularly relevant to the current study are the growing number of studies that have applied the framework to pedagogical interventions designed to enhance the connection between the ideal L2 self and motivated behavior by expanding, for example, the vividness of learners' ideal self and thus increasing learners' motivational capacity (Dörnyei & Kubanyiova, 2014; Magid & Chan, 2012; Safdari, 2021; Sato, 2021; Sato & Lara, 2019). A pioneering study by Magid and Chan (2012) reported on two different intervention programs in England and Hong Kong. Learners participated in activities such as drawing a timeline, developing an action plan with specific steps, and clarifying their vision of feared selves. These interventions led to stronger visions of learners' ideal L2 selves and increased confidence and effort towards learning English. Magid and Chan's program in Hong Kong was based on the *Possible Selves Program*, originally developed in the field of education (Hock et al., 2006). In Hock et al.'s (2006) original study, it was found that students who participated in the program identified a larger number of goals, and that their goals were articulated with more specificity than their peers. Especially relevant to the current study is the emphasis within the interventions on articulating action plans and goals.

More recently, studies such as Safdari (2021), Sato (2021), and Sato and Lara (2019) have also implemented vision enhancement studies in EFL contexts, applying the six major steps proposed by Dörnyei and Kubanyiova (2014). They provide support for the positive effects of vision-centered pedagogical interventions on aspects of motivation such as learners' visions of their ideal L2 self (Sato, 2021; Sato & Lara, 2019), intended effort, and learners' actual target language use (Sato, 2021). Although these studies support the effectiveness of vision enhancement, some questions still remain. In particular, Thorsen et al. (2020) argue for

the need to focus on one of the key driving forces of the L2MSS model, the discrepancy between the current and L2 self. They propose that change in motivation is a function of changes in learners' understanding of their current L2 self as well as their future guides. This point may be especially relevant for helping learners to actually engage in motivated behavior by "transforming the vision into action," and corresponds to the step "providing students with self-relevant roadmaps" (Dörnyei & Kubanyiova, 2014, p.101). Although measures of the current L2 self for research purposes have been included in some previous studies (Henry & Cliffordson, 2017; MacIntyre et al., 2009), the current study proposes that, for pedagogical contexts, standardized assessment may serve as a useful guide to "the currently missing current L2 self" (Thorsen et al., 2020, p. 597).

The Current Study

The pedagogical intervention for the current study is an assessment cycle, consisting of self-assessment, standardized assessment, assessment feedback, and goal setting. It was designed to harness the widely applicable and easily implemented classroom practice of assessment as a motivational strategy, which, according to the L2MSS framework could help learners to understand the distance between their current state and their ideal state, and thus help learners to plan and put into action the steps they would need to take to reach their ideal state. The current study focuses on motivation for developing one specific L2 skill, speaking, and takes a longitudinal approach to investigating the impact of the pedagogical intervention by tracking motivational dynamics over the course of seven months (Campbell & Storch, 2011; Waninge et al., 2014). The following research questions are addressed:

RQ 1. Are there any changes in quantitative measures of motivation and motivated learning behavior of EFL learners who participate in a pedagogical intervention?

RQ 2. Are there any qualitative changes in motivated learning behavior of EFL learners who participate in a pedagogical intervention?

Method

Participants

A total of 78 university students participated in the current study. They consisted of 50 first year, 17 second year, and 19 third year students, majoring in English at two women's universities in Tokyo, Japan. The universi-

ties were comparable in size, range of English proficiency, and curriculum for English majors. The learners, all women, had received six years of formal English education at junior and senior high school, and their English proficiency level was considered to range from low-intermediate to high-intermediate levels based on placement tests at their universities. They were recruited in their English courses. Intervention was conducted with a sub-group of the learners who belonged to two intact classes (*n*=21), each taught by one of the researchers, referred to in the study as Group 1. Learners who did not belong to these two classes did not participate in the intervention (*n*=57) and are referred to in the study as Group 2.

Because the data was collected in a natural context, it was not possible to control for content of learners' English classes and variation in learners' selection of elective English classes. All learners were taking one or more English classes typical of English majors in their universities.

Figure 1

Overview of Research Design

	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
Group 1 (*n*=21)	Intervention 1 →					Intervention 2 →	
	Motivation Measure (7 times) Self-Assessment (7 times) Motivated Learning Behavior Measure (7 times)						
Group 2 (*n*=57)	No Intervention						
	Motivation Measure (3 times) Self-Assessment (3 times) Motivated Learning Behavior Measure (4 times)						

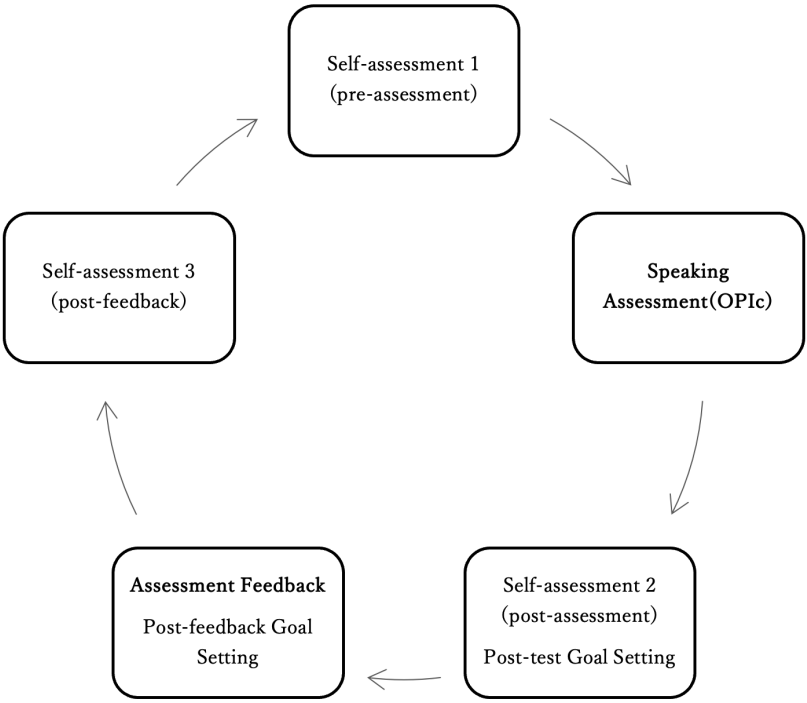
Design

The study is a descriptive and longitudinal study that investigates the trajectories of motivational variables of two groups of learners over a period of seven months. As shown in Figure 1, learners in Group 1 participated in a pedagogical intervention and learners in Group 2 did not. Group 1 experienced two cycles of the same assessment and feedback intervention. Both groups completed measures of motivation, self-assessment, and motivated learning behavior.

Pedagogical Intervention: Speaking Assessment and Feedback Cycle

The pedagogical intervention, which aimed to provide learners with the tools to put their visions into action, consisted of four components: (a) a standardized speaking assessment, (b) a self-assessment, (c) feedback on the standardized speaking assessment, and (d) goal setting, which were arranged in a cycle as shown in Figure 2.

Figure 2
Intervention Cycle



a. Standardized speaking assessment: The ACTFL Oral Proficiency Interview-computer (OPIc) was used as a speaking assessment. The

rating scale for the OPIc assigns eight proficiency levels: low novice, mid novice, high novice, low intermediate, mid intermediate (1, 2), high intermediate, and low advanced.

b. Self-assessment: The self-assessment tool, which learners completed three times during one intervention cycle, asked learners to assess their own speaking ability and consisted of 30 can-do statements taken from the Swiss version of the European Language Portfolio (Little & Perclova, 2001). Learners evaluated their ability to do what was described in each statement using a 4-point Likert scale. Items included statements such as "I can introduce myself," and "I can buy tickets and ride public transportation."
c. Feedback on the standardized speaking assessment: Learners received official reports of their rating approximately two weeks after administration of the oral proficiency test. The official report included a description of the relevant proficiency level, which was supplemented by explanation of the relevant proficiency levels in the learners' native language, Japanese, prepared for the purpose of the study. The performance-level descriptors for the ACTFL Oral Proficiency Interview specify in detail learners' language skills and knowledge at each level.
d. Goal-setting worksheets: Learners completed a post-test goal-setting worksheet and post-feedback goal-setting worksheet after each administration of the oral proficiency test in their native language, Japanese. The post-test goal-setting worksheet asked learners to record their reactions to the oral proficiency test immediately after taking the test. Learners were asked to describe (a) concrete goals for improving their performance on the next test, and (b) how they should study in order to improve their performance on the next test. The post-feedback goal-setting worksheet asked learners to record their reactions after receiving feedback on the test. Learners were asked to describe (a) what they needed to improve, (b) what they needed to do to achieve their goal, and (c) how they should prepare for the next test.

Materials

The three following measures were used to chart the motivational trajectories of learners in both Group 1 and Group 2.

Motivation

Learners filled out a motivation questionnaire multiple times throughout the course of the study: seven times for learners in Group 1 and three

times for learners in Group 2. The questionnaire contained 10 question items adapted from questionnaires used in studies by Yashima (2002) and Gardner (2010). The items focused on learners' desire to improve their motivation, such as "I want to improve my speaking," as well as learners' motivational effort, such as "I make an effort to improve my speaking," and "I think I spend fairly long hours studying English." Learners responded on a 7-point Likert scale. See Appendix A for the complete questionnaire.

Self-Assessment of Speaking Ability

The self-assessment measure, described above as part of the intervention cycle also served as a measure of learners' perceptions of their own speaking ability. The self-assessment measure was completed at seven different time points by Group 1 as part of the pedagogical intervention and at three time points by Group 2.

Table 1
Data Collection Procedure

Time points	Group 1	Group 2 (no intervention)
	Measures	Measures
1. Pre-test 1 Before 1st intervention (June)	Motivation 1 Self-assessment 1 Learning behavior 1	Motivation 1 Self-assessment 1 Learning behavior 1
2. Post-test 1 After 1st standardized speaking assessment (June)	Motivation 2 Self-assessment 2	
3. Post-feedback 1 After feedback from standardized speaking assessment (July)	Motivation 3 Self-assessment 3 Learning behavior 2	Learning behavior 2
4. After summer Between interventions (September)	Motivation 4 Self-assessment 4 Learning behavior 3	Motivation 2 Self-assessment 2 Learning behavior 3

5. Pre-test 2 Before 2nd intervention (December)	Motivation 5 Self-assessment 5 Learning behavior 4	Motivation 3 Self-assessment 3 Learning behavior 4
6. Post-Test 2 After 2nd standardized speaking assessment (December)	Motivation 6 Self-assessment 6	
7. Post-Feedback 2 After feedback from standardized speaking assessment (January)	Motivation 7 Self-assessment 7 Learning behavior 5	

Motivated Learning Behavior

Learners also filled out a second questionnaire which asked them to report on their motivated learning behavior during the previous month at multiple time points: five different time points for learners in Group 1 and four time points for learners in Group 2. Learners were asked (a) how much time (in hours and minutes) they had spent to improve their speaking skills outside the classroom per day, and (b) what they had actually done during class time.

Procedure

The study was conducted over a period of seven months. The study was approved by an institutional research ethics committee. Informed consent was obtained from participants at the beginning of the study. As described above, Group 1 participated in two intervention cycles, once in June/July and once in December/ January. As shown in Table 1, Group 1 completed motivation questionnaires and self-assessments at seven time points, and reported on their motivated learning behavior at five time points. Learners in Group 2 completed motivation questionnaires and self-assessment at three time points, and reported on their learning behavior at three time points.

Data and Analysis

The aim of the current study was to examine the trajectory of learner motivation over time with two groups of learners. Data was collected at multiple time points. Therefore, the independent variables in the current

study were Group and Time Points. Dependent variables were motivation, self-assessments of speaking ability, and motivated learning behavior. All of the statistical analyses were performed using the statistical software SPSS 24.0.

Data on Dependent Variables

Motivation. Data consisted of responses to 10 items on a 7-point Likert scale. Possible total scores ranged from a minimum of 10 to a maximum of 70 points. The result of the factor analysis for motivation (see Appendix B) yielded the anticipated two factors: Desire to Improve Speaking and Motivational Intensity. Each of them obtained appreciable loadings (i.e., loadings of more than .35) from the corresponding items. This factor structure supported the presupposition that these two subscales assessed different components of motivation, namely the elements of desire and effort, both of which should be included in an index of motivation according to Gardner (2010). The Cronbach alpha indices of the subscales were .74 and .78, indicating that the items in each subscale had an adequate level of internal consistency (George & Mallery, 2003). Therefore, the total score of motivation was calculated as the sum of the two subscales on motivation.

Self-Assessment. Data consisted of responses to 30 items on a 4-point Likert scale. The possible total scores ranged from 30 to 120 points. Factor analysis yielded three factors: Level A, Level B, and Level C (see Appendix B), each of which obtained considerable loadings from the items corresponding to one of the three proficiency levels (i.e., A, B, and C) categorized in CEFR, illustrating that these subscales successfully served as a can-do list, tapping skills at different difficulty levels. The Cronbach alpha indices were .82, .90, and .80, demonstrating the internal consistency of these subscales.

Motivated Learning Behavior. The data for motivated learning behavior consisted of learners' reports of the average amount of time per day in hours and minutes they spent on improving their speaking skills in the previous month and learners' reports about the specific content or type of motivated behavior. Using grounded analysis, learners' comments were examined for salient themes, and emergent categories for types of learning behavior were identified. After socialization and agreement on the coding categories by both researchers, learner's comments on the remainder of the dataset were then coded by one of the researchers based on these emergent categories. A total of 20% of the data was also coded independently by the other researcher. Interrater-reliability was very high as indicated by Cohen's Kappa ($\kappa = 0.821$).

Data on Intervention

Standardized Speaking Assessment. Table 2 shows learners' official ACTFL OPIc ratings for the first administration in June and second administration in December. In the first administration, eight learners were assigned to low intermediate, making it the most commonly assigned level. In the second administration, five learners were assigned to low intermediate and six learners were assigned to mid intermediate 1. In terms of individual improvement between the first and second administration of the OPIc, eight learners improved at least one level, nine learners maintained the same level, and five learners were assigned a lower level than their previous rating.

Table 2

Number of Learners Assigned to Each OPIc Proficiency Level (n = 21)

	Novice High	Low Intermediate	Mid Intermediate 1	Mid Intermediate 2	High Intermediate	Advanced
June	2	8	5	2	3	1
December	3	5	6	2	4	1

Goal-Setting. Learners' goals elicited on their worksheets during the assessment interventions were reported in a previous study (Fujii, 2018), and reflected four types of goals: (1) opportunities for speaking practice, (2) ability to articulate intended meaning in English, (3) knowledge of vocabulary, and (4) knowledge of content (see Fujii, 2018 for detailed descriptions and examples of each category).

Results

Research Question 1

In order to address the first research question *Are there any changes in quantitative measures of motivation and motivated learning behavior of EFL learners who participate in a pedagogical intervention?* multivariate analysis of variance (MANOVA) was first conducted with Group as the between-subjects factor and Time Point as the within-subjects factor to

examine the change in the trajectories of motivation, self-assessment, and motivated learning behavior across time for both Group 1 and Group 2. In preliminary tests, no significant violation of the univariate normality assumption was identified, with skewness and kurtosis both within 3.29. Furthermore, no multivariate outliers were detected at the significance level of .001, as assessed by Mahalanobis distances (Tabachnick & Fidell, 2007). In addition, the result of the Box's M test confirmed the equality of covariance matrices of dependent variables despite the sample size difference of the two groups ($p = .25$).

The results of the MANOVA, according to Pillai's trace, indicated that while the main effect of Group was not significant, $F(3, 74) = 1.89$, $p = .14$, partial eta squared = .07, that of Time Point was significant, $F(6, 71) = 4.85$, $p < .001$, partial eta squared = .29. In addition, a significant interaction between Group and Time Point on dependent variables was observed, $F(6, 71) = 6.51$, $p < .001$, partial eta squared = .36. Therefore, the simple main effect of each variable was next examined using a univariate analysis of variance (ANOVA).

The results of the univariate ANOVA are reported in Table 3 and show that the interaction effect of Group and Time Point was significant for both motivation ($F(1.79, 136.29) = 4.53$, $p < .016$, partial eta squared = .056) and self-assessment *(*$F(1.68, 127.28) = 12.62$, $p < .001$, partial eta squared = .14). This means that for motivation and self-assessment, there were differences in how Group 1 and 2 changed across time. The trajectories of each variable are presented in more detail below. In running the ANOVA, Greenhouse-Geisser values were used for both motivation and self-assessment, in order to correct for violation of sphericity. Bonferroni adjustment of the p-values was employed ($p < .016$) in order to counteract the problem of multiple comparisons.

Table 3
Results of Univariate ANOVA

Source	*DV*	SS	df	*MS*	*F*	*p*	η_p^2
Between Subjects							
Group	M	139.01	1	139.01	1.56	0.23	0.02
	SA	2317.69	1	2317.69	3.80	0.06	0.05
	LB	1049.46	1	1049.46	0.75	0.39	0.01
Error	M	6794.32	76	89.40	-	-	-
	SA	46351.21	76	609.88	-	-	-
	LB	106845.95	76	1405.87	-	-	-
Within Subjects							
Time Point	M	122.20	1.79	68.14	4.31	0.02	0.05
	SA	413.90	1.68	247.15	4.01	0.03	0.05
	LB	1142.04	2	571.02	3.37	0.04	0.04
Time Point × Group	M	128.49	1.79	71.65	4.53	0.015*	0.06
	SA	1304.32	1.68	778.84	12.62	0.000*	0.14
	LB	221.95	2	110.98	0.65	0.52	0.01
Error (Time Point)	M	2155.56	136.29	15.82	-	-	-
	SA	7854.44	127.28	61.71	-	-	-
	LB	25795.78	152	169.71	-	-	-

Note. M: motivation, LB: learning behavior, SA: self-assessment. * indicates *p* value < .016.

Motivation

Table 4 and Figure 3 show the means and standard deviations of motivation of Group 1 at seven time points of data collection and Group 2 at three time points. The results of the ANOVA indicated a significant interaction between group and time *(F*(1.79, 136.29) = 4.53, $p < .016$), partial eta squared = .06. As can be seen in the graph, the motivation of both groups declined

similarly between Pre-Test 1 through After Summer, which includes the period before, during, and after the first intervention.

Table 4

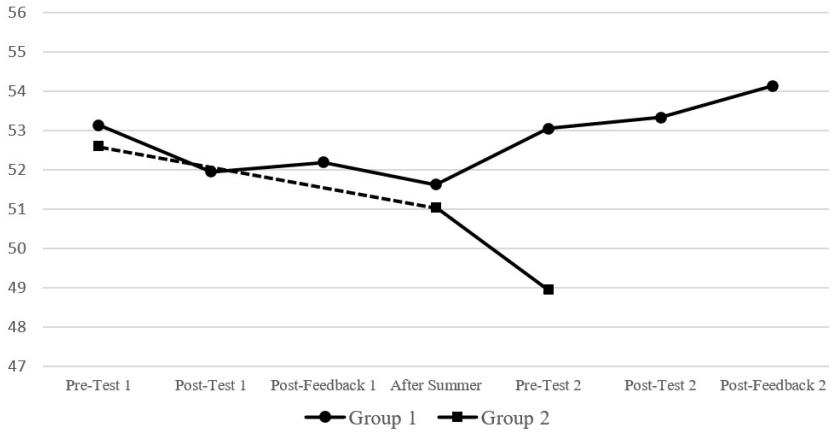
Means and Standard Deviations for Motivation

	Group 1		Group 2	
Time Points	*M*	*SD*	*M*	*SD*
1. Pre-Test 1	53.14	6.48	52.60	6.25
2. Post-Test 1	51.95	6.82	-	-
3. Post-Feedback 1	52.19	7.33	-	-
4. After Summer	51.62	6.41	51.05	6.74
5. Pre-Test 2	53.05	5.63	48.95	5.86
6. Post-Test 2	53.33	6.00	-	-
7. Post-Feedback 2	54.14	5.62	-	-

Note. n = 21 (Group 1), 57 (Group 2). Possible score range: 10-70.

Figure 3

Changes in Motivation



Note. A dotted line indicates data points for Group 2 which are further apart than the data points for Group 1.

However, a noticeable difference between the two groups appeared at Pre-Test 2, right before the second pedagogical intervention cycle. At this point, the motivation of Group 2 clearly dropped to a lower level than that at Pre-Test 1, whereas the motivation level of Group 1 was largely maintained during the seven time points. Post-hoc tests confirmed that the gap between the two groups resulted in a statistically significant difference at Pre-Test 2 with a relatively large effect size, $t(76) = 2.77$, $p < .01$, Cohen's $d = .71$.

Self-Assessment

Table 5 and Figure 4 show means and standard deviations for the self-assessment scores of Group 1 at seven time points of data collection and Group 2 at three time points.

Table 5

Means and Standard Deviations for Self-Assessment

	Group 1		Group 2	
Time points	*M*	*SD*	*M*	*SD*
1. Pre-test 1	71.33	14.14	70.91	15.23
2. Post-test 1	66.33	17.69	-	-
3. Post-feedback 1	52.84	15.22	-	-
4. After summer	62.48	14.39	73.68	15.46
5. Pre-test 2	62.57	15.41	73.07	16.34
6. Post-test 2	67.95	18.66	-	-
7. Post-feedback 2	72.85	16.61	-	-

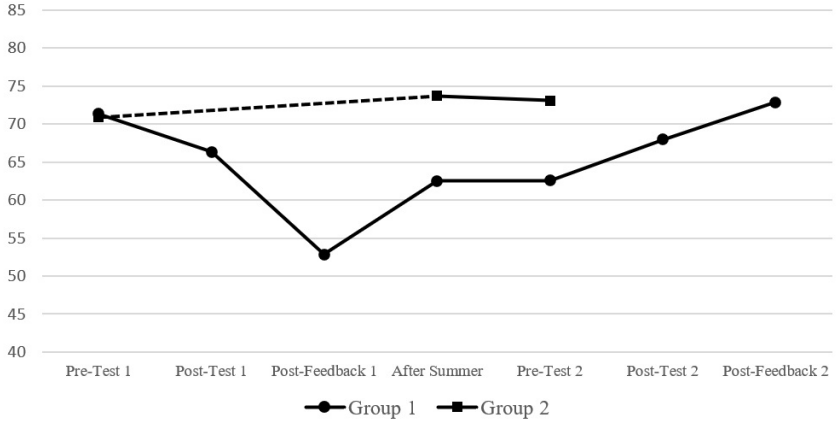
Note. $n = 21$ (Group 1), 57 (Group 2). Possible score range: 30-120.

With respect to self-assessment, the results of the ANOVA indicated a significant interaction between group and time, $F(1.68, 127.28) = 12.62$, $p < .001$, partial eta squared = .14. Post-hoc tests confirmed the trends visible in Figure 4. First, they indicated that the self-assessment scores for Group 1 at Post-Feedback 1 were significantly lower than any other time period ($p < .001$), $t(20) = 6.34 \sim 9.79$, $p < .001$, Cohen's $d = .64 \sim 1.26$, and also that Group 1 scores were noticeably lower than that of Group 2 even at the two subsequent time points where Group 2 completed self-assessment: After

Summer, $t(76) = -2.89$, $p < .01$, Cohen's $d = .74$, and Pre-Test 2, $t(76) = -2.56$, $p < .016$, Cohen's $d = .65$.

Figure 4

Changes in Self-Assessment



Note. A dotted line indicates data points for Group 2 which are further apart than the data points for Group 1.

In short, although the self-assessment of Group 2 was relatively stable, that of Group 1 dropped significantly after they received the feedback of their first speaking test. The self-assessment of Group 1 recovered after the summer break, but still stayed significantly lower than that of the control group after summer until gradually rising again after the second standardized speaking assessment.

Motivated Learning Behavior

Table 6

Means and Standard Deviations for Motivated Learning Behavior (Minutes per Day)

	Group 1		Group 2	
Time points	*M*	*SD*	*M*	*SD*
1. Pre-test 1	30.71	25.75	28.77	23.27
2. Post-test 1	-	-	-	-
3. Post-feedback 1	31.90	27.68	28.60	26.67
4. After summer	38.33	25.61	33.25	23.93
5. Pre-test 2	35.71	25.11	28.42	23.63
6. Post-test 2	-	-	-	-
7. Post-feedback 2	35.48	24.59	-	-

Note. n = 21 (Group 1), 57 (Group 2).

Figure 5

Changes in Motivated Learning Behavior (Amount of Time Spent)

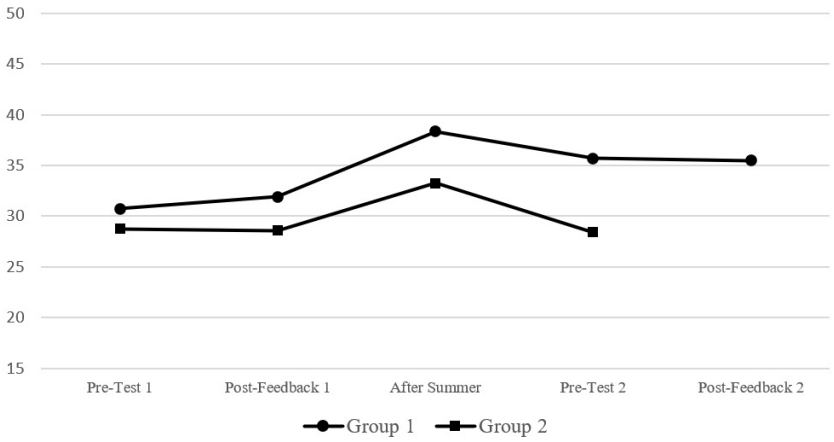


Table 6 and Figure 5 show the means and standard deviations for motivated learning behavior of Group 1 at five time points of data collection and Group 2 at four time points. In comparing the learning behavior (amount of time spent) of the two groups over the course of seven months, Figure 5 shows that Group 1 generally spent more time in speaking practice than Group 2. Nevertheless, as reported above and shown in Table 3, the result of the univariate ANOVA revealed that there was no statistically significant difference in motivated learning behavior between the two groups or between the different time points.

Research Question 2

This section addresses the second research question, *Are there any qualitative changes in motivated learning behavior of EFL learners who participate in a pedagogical intervention?*

Table 7

Number of Reports on Types of Motivated Learning Behavior

	Total reports				Reports per learner			
	Pre-test 1	Post-feedback 1	After summer	Pre-test 2	Pre-test 1	Post-feedback 1	After summer	Pre-test 2
Group 1	31	25	20	21	1.48	1.19	0.95	1
Group 2	34	44	29	29	0.6	0.77	0.51	0.51

Note. n = 21 (Group 1), 57 (Group 2).

Table 7 shows the number of learners' reports about the specific content of their motivated learning behavior at four points in time. These descriptions of motivated learning behavior were categorized into six categories that emerged through qualitative analysis: general output, specific output, general input, specific input, pronunciation, and vocabulary, as shown in Table 8.

Table 8
Coding Categories and Examples for Learners' Motivated Learning Behavior

Category and description	Examples
General output: General reference to quantity of output.	*I tried to participate in class; I spoke to my classmate in English.*
Specific output: Specific references to nature of output.	*I tried to use simple sentences to express my opinions; I tried using many different grammar forms.*
General input: General reference to value placed on input.	*I listened carefully to my classmates' English.*
Specific input: Specific reference to ability to comprehend input or how input was used.	*I clarified the meaning when I couldn't understand; I input the phrases my teacher used.*
Pronunciation: Reference to pronunciation.	*I listened to the native speaker teacher's pronunciation; I was careful of pronunciation when I read aloud.*
Vocabulary: Reference to vocabulary which is unrelated to input or output	*I wrote down unfamiliar words.*

Tables 9 and 10 show the number of learners' reports in each of the six categories for both Groups 1 and 2 as a percentage of the total number of reports. For Group 1, there was a decline in general comments about output from 48% of all reports in June (before the speaking assessment) to 36% of all reports in July (before summer vacation) and an increase in specific comments about output during this same period from 6% to 20% of all reports as shown in Figure 6. In July, one month after the intervention, learners reported engaging in behavior that was described in more specific terms such as "I spoke without looking at my notes," "I tried to paraphrase so that my English is easier to understand," or "I tried to incorporate more filler expressions."

Table 9
Types of Learning Behavior Reported by Group 1 (Percentage)

	June	July	Oct	Dec
General reference to quantity of output	48%(15)	36% (9)	30% (6)	33% (7)
Specific reference to quality of output	6% (2)	20% (5)	20% (4)	24% (5)
General reference to quantity of input	13% (4)	4% (1)	5% (1)	14% (3)
Specific reference to quality of input comprehension	16% (5)	24% (6)	15% (3)	5% (1)
Pronunciation	13% (4)	12% (3)	15% (3)	14% (3)
Reference to vocabulary unrelated to input or output	3% (1)	4% (1)	15% (3)	10% (2)

Note. Raw numbers are shown in ( ). Percentage points have been rounded.

Similarly, general reports about input declined 9% between June and July, while specific comments about input increased 8% during this time as shown in Table 9 and Figure 7. In July, learners reported more focused behavior such as "asked questions to clarify what my classmates meant," "when listening to classmates' presentations, compared the language to what I had prepared." The increase in specific comments was not maintained between October and December.

Table 10
Types of Learning Behavior Reported by Group 2 (Percentage)

	June	July	Oct	Dec
General reference to quantity of output	53% (18)	50% (22)	59% (17)	28% (8)
Specific reference to quality of output	12% (4)	5% (2)	7% (2)	17% (5)
General reference to quantity of input	0% (0)	2% (1)	7% (2)	7% (2)
Specific reference to quality of input comprehension	15% (5)	9% (4)	3% (1)	14% (4)
Pronunciation	18% (6)	30% (13)	14% (4)	31% (9)
Reference to vocabulary unrelated to input or output	3% (1)	5% (2)	10% (3)	3% (1)

Note. Raw numbers are shown in ( ). Percentage points have been rounded.

Figure 6
Changes in Learning Behavior of Group 1 (Types of Behavior): Output

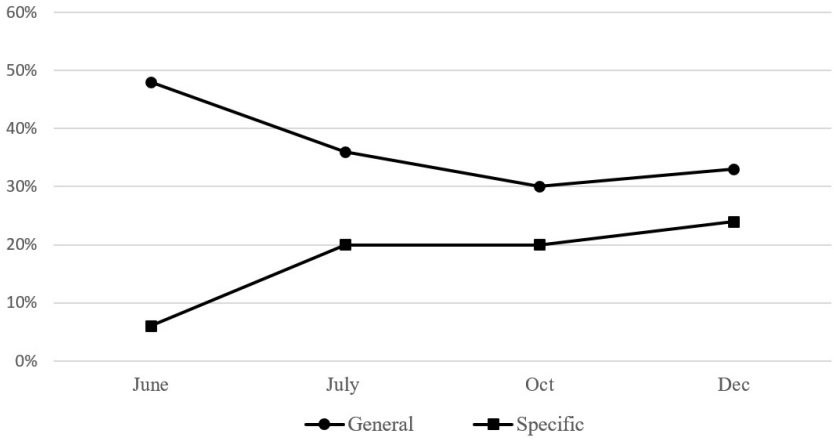
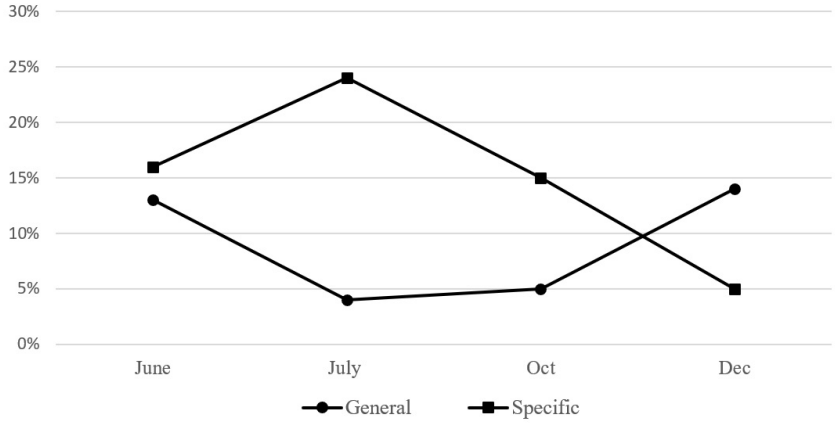


Figure 7

Changes in Learning Behavior of Group 1 (Types of Behavior): Input



In contrast, for Group 2, general reports about output remain high from June through to October, and there was no increase in specific comments about output or input in July or October, although there was an increase in specific comments about both output and input between October and December as shown in Table 10. Thus, while the data indicated no effect of the pedagogical intervention on learning behavior in terms of amount of effort, learners' reports of their learning behavior showed an increase in focused behavior related to both input and output particularly for Group 1.

Discussion and Conclusions

The goal of the current study was to explore the effectiveness of a pedagogical intervention that included speaking assessment, assessment feedback, and goal setting on learners' motivation and motivated learning behavior for improving their speaking skills. The findings of the study highlight interesting trends in the quantitative and qualitative trajectories of learners' motivation in the group that participated in the speaking assessments, especially in comparison to the group which did not participate in the intervention.

First, as displayed in Figure 8, the learners in the intervention group (Group 1) maintained their motivation over the course of time whereas the

non-intervention group (Group 2) dropped in motivation. Given that previous studies have indicated that motivation generally tends to decline over time (Thorson et al., 2020), the trends for Group 2, align with the default trajectory, while Group 1 displayed a marked path, possibly attributable to having experienced the assessment cycle.

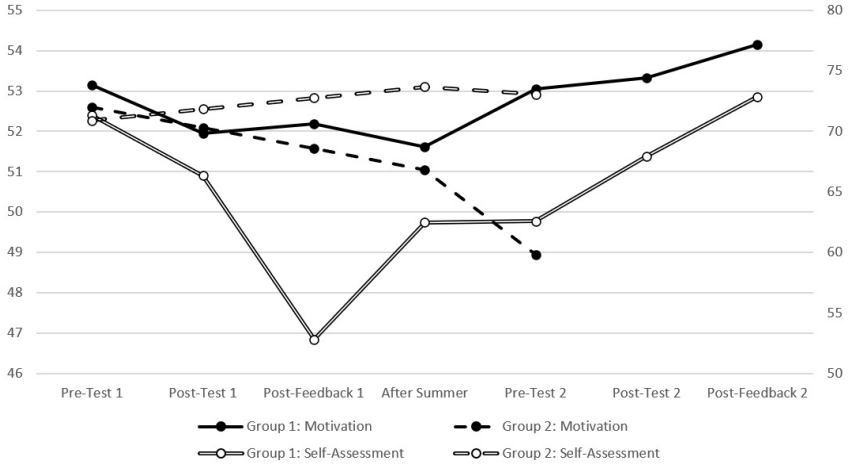
Second, one of the most interesting findings was the significant drop in learners' self-assessment scores after the pedagogical intervention, specifically after the feedback session. This change in self-perception suggests that pedagogical intervention was effective in impacting learners' understanding and evaluation of their current state. Also noteworthy is the timing in the drop of the self-assessment scores after receiving the test feedback rather than immediately after taking the speaking test, suggesting that their perceptions of their own speaking ability were not necessarily modified by the actual test experience, but through the feedback session which also included a goal-setting component. Because the speaking assessment assigned learners to bands with clear descriptors of their performance characteristics, it is likely that learners were able to understand their current speaking ability in objective and comprehensible terms as well as the gap between their current level and the next level, which mostly likely led to the clarification of their immediate learning goals. In other words, learners' increased awareness of their current state most likely provided them with a reliable picture of the discrepancy between their current and future self, a tension which is key in driving motivated behavior (Henry & Cliffordson, 2017; Macintyre et al., 2009; Thorson et al., 2020) and may have led to a clearer view of the procedures for reaching their goal (or ideal L2 self). Conversely, the non-intervention group maintained a higher evaluation on their self-assessment, most likely because they did not have such experiences.

Finally, the findings of the study indicated that the assessment intervention did not impact the quantity of learners' motivated learning behavior (how much they studied), but did impact the quality of their motivated learning behavior (how they studied), which became more focused after the assessment. This serves as some evidence that the assessment made an impact on the learners' roadmaps to their goal.

Thus, taken together, these findings align with the proposed role of the assessment intervention in helping learners to understand their current L2 self, which in turn helped learners to revise their action plans for achieving their goals.

Figure 8

Visual Summary of Learners' Motivation and Self-Assessment Over Time



Note. The x-axis shows the timing of tests and interventions over the 7-month span. The y-axis on the left shows the level of learners' motivation. The y-axis on the right represents the level of learners' self-assessment. The lines on the graph represent the motivation and self-assessment for Groups 1 and 2 over time.

It should be noted, however, that over the long term, the interplay in dynamics may be more complex. As illustrated in Figure 8, the motivation and self-assessment of Group 1, the intervention group, showed different moves after the first and second tests. Their motivation went down after the first test but marginally increased after the second test. Similarly, whereas their self-assessment fell dramatically after the first test, it slightly improved after the second one. An assessment intervention may have a different impact on motivation and self-assessment depending on the timing or frequency of administration. More longitudinal research is needed to further understand the complexity of the motivational system, and educators should keep this in mind when they incorporate assessment practices in their L2 teaching.

In sum, the findings support a view of motivation which connects an assessment intervention, motivation, and motivated learning behavior, and where assessment as a motivational strategy worked to enhance motiva-

tion and transform motivation to motivated learning behavior. The results of the study suggest that learners may benefit from pedagogical practices that help learners to comprehend their current L2 self and develop concrete and realistic steps that are necessary for realizing their goals. Assessment can be one useful motivational strategy when accompanied by feedback and opportunities for reflection.

Limitations of the study must be recognized. First, because of the ecological context of the study, the two groups were not matched. Some of the differences that were observed over the course of the study may be attributable to factors other than the pedagogical intervention. Thus, there may have been additional factors that contributed to the downward trend in motivation of Group 2. Also, because the learners represent learners and learning in a specific context, the findings of the study may not be applicable to other learners in other contexts. There may have been external factors specific to these groups of learners such as extra-curricular priorities and career goals, for example, that restricted the quantity of time learners directed towards speaking practice outside of the classroom. Finally, whether the benefits of the pedagogical intervention impact learning achievement remains an important question for future study. Although the results of the proficiency tests did not show overall improvement during the period of the current study, it could be that more fine-grained measures of proficiency and a more extended period of study may be necessary to track meaningful change. Further investigation into refining the measures of motivated learning behavior and feedback and goal setting stages of the intervention may also shed more light on the relationship between assessment and improvements in quantity and quality of motivated behavior, and ultimately gains in speaking proficiency.

Acknowledgments

This research study was supported by Japan Society for the Promotion of Science (JSPS) Grant-in-Aid for Scientific Research (C) Grant Number 26370738. The authors would like to thank the participants and research assistants for their contribution to the project.

Akiko Fujii is an associate professor at the International Christian University in Tokyo. Her research interests focus on English language learning in the Japanese context, especially with relation to task-based language teaching, classroom interaction, English-medium instruction, and teacher development.

Yoshinori Inagaki is a professor at Tsuda University. His main research interests include learner factors in second language acquisition, particularly motivation and learning strategies, as well as English education at the elementary school level in Japan.

Appendices

All appendices are available from the online version of this article at https://jalt.org/main/jj.

References

Boo, Z., Dörnyei, Z., & Ryan, S. (2015). L2 motivation research 2005-2014: Understanding a publication surge and a changing landscape. *System*, *55*, 145–157. https://doi.org/10.1016/j.system.2015.10.006

Campbell, E., & Storch, N. (2011). The changing face of motivation: A study of second language learners' motivation over time. *Australian Review of Applied Linguistics*, *34*(2), 166–192. https://doi.org/10.1075/aral.34.2.03cam

Carless, D. (2015). Exploring learning-oriented assessment processes. *Higher Education*, *69*(6), 963–976. https://doi.org/10.1007/s10734-014-9816-z

Cheng, H., & Dörnyei, Z. (2007). The use of motivational strategies in language instruction: The case of EFL teaching in Taiwan. *Innovation in Language Learning and Teaching*, *1*(1), 153–174. https://doi.org/10.2167/illt048.0

Chong, S.W., & Reinders, H. (2023). *Innovation in learning-oriented language assessment*. Palgrave Macmillan. https://doi.org/10.1007/978-3-031-18950-0

Csizér, K. (2019). The L2 motivational self system. In M. Lamb (Ed.), *The Palgrave handbook of motivation for language learning* (pp. 71–93). Palgrave Macmillan. https://doi.org/10.1007/978-3-030-28380-3

Csizér, K., & Kormos, J. (2009). Modelling the role of inter-cultural contact in the motivation of learning English as a foreign language. *Applied Linguistics*, *30*(2), 166–185. https://doi.org/10.1093/applin/amn025

Dörnyei, Z. (2001). *Motivational strategies in the language classroom.* Cambridge University Press. https://doi.org/10.1017/CBO9780511667343

Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. Lawrence Erlbaum. https://doi.org/10.1075/aila.19.05dor

Dörnyei, Z. (2009). The L2 motivational self-system. In Z. Dörnyei & E. Ushioda (Eds.), *Motivation, language identity, and the L2 self* (pp. 9–42). Multilingual Matters. https://doi.org/10.21832/9781847691293-003

Dörnyei, Z., & Csizér, K. (1998). Ten commandments for motivating language learners: Results of an empirical study. *Language Teaching Research*, *2*(3), 203–229. https://doi.org/10.1177/136216889800200303

Dörnyei, Z., & Kubanyiova, M. (2014). *Motivating learners, motivating teachers: Building vision in the language classroom.* Cambridge University Press.

Fox, J. (2014). Large-scale assessment. In A. J. Kunnan (Ed.), *The companion to language assessment: Approaches and development* (pp. 533–541). Wiley-Blackwell. https://doi.org/10.1002/9781118411360.wbcla022

Fujii, A. (2018). The role of speaking assessment in shaping EFL learners' subsequent learning goals. *Educational Studies, 60*, 61–72.

Gan, Z., Leung, C., He, J., & Nang, H. (2019). Classroom assessment practices and learning motivation: A case study of Chinese EFL students. *TESOL Quarterly*, *53*(2), 514–529. https://doi.org/10.1002/tesq.476

Gardner, R. C. (2010). *Motivation and second language acquisition: The socio-educational model.* Peter Lang Publishing.

Gebril, A. (2023). *Learning-oriented assessment: Putting theory into practice*. Routledge.

Guilloteaux, M. J., & Dörnyei, Z. (2008). Motivating language learners: A classroom-oriented investigation of the effects of motivational strategies on student motivation. *TESOL Quarterly*, *42*(1), 55–77. https://doi.org/10.1002/j.1545-7249.2008.tb00207.x

Henry, A., & Cliffordson, C. (2017). The impact of out-of-school factors on motivation to learn English: Self-discrepancies, beliefs, and experiences of self-authenticity. *Applied Linguistics*, *38*(5), 713–736. https://doi.org/10.1093/applin/amv060

Hiver, P., & Al-Hoorie, A. H. (2020). Reexamining the role of vision in second language motivation: A preregistered conceptual replication of You, Dörnyei, and Csizer (2016). *Language Learning*, *70*(1), 48–102. https://doi.org/10.1111/lang.12371

Hock, M. F., Deshler, D. D., & Schumaker, J. B. (2006). Enhancing student motivation through the pursuit of possible selves. In C. Dunkel & J. Kerpelman (Eds.), *Possible selves: Theory, research and application* (pp. 205–221). Nova Science.

Jang, E. E., & Wanger, M. (2014). Diagnostic feedback in the classroom. In A. J. Kunnan (Ed.), *The companion to language assessment: Approaches and development* (pp. 693–711). Wiley-Blackwell. https://doi.org/10.1002/9781118411360.wbcla081

Kissling, E. M., & O'Donnell, M. E. (2015). Increasing language awareness and self-efficacy of FL students using self-assessment and the ACTFL proficiency guidelines. *Language Awareness*, *24*(4), 283–302. https://doi.org/10.1080/09658416.2015.1099659

Lamb, M. (2012). A self system perspective on young adolescents' motivation to learn English in urban and rural settings. *Language Learning*, *62*(4), 997–1023. https://doi.org/10.1111/j.1467-9922.2012.00719.x

Lamb, M. (2017). The motivational dimension of language teaching. *Language Teaching*, *50*(3), 301–346. https://doi.org/10.1017/S0261444817000088

Lamb, M. (2019). Motivational teaching strategies. In M. Lamb, K. Csizer, A. Henry, & S. Ryan (Eds.), *The Palgrave handbook of motivation for language learning*, (pp. 287–305). Palgrave Macmillan. https://doi.org/10.1007/978-3-030-28380-3

Le-Thi, D., Dörnyei, Z., & Pellicer-Sánchez, A. (2022). Increasing the effectiveness of teaching L2 formulaic sequences through motivational strategies and mental imagery: A classroom experiment. *Language Teaching Research, 26*, 1202–1230. https://doi.org/10.1177/1362168820913125

Little, D., & Perclova, R. (2001). *The European language portfolio: A guide for teachers and teacher trainers*. Council of Europe. https://www.coe.int/en/web/portfolio/elp-related-publications

MacIntyre, P. D., Mackinnon, S. P., & Clement, R. (2009). Toward the development of a scale to assess possible selves as a source of language learning motivation. In Z. Dörnyei & E. Ushioda (Eds.), *Motivation, language identity and the L2 self* (pp. 193–214). Multilingual Matters. https://doi.org/10.21832/9781847691293-011

Magid, M., & Chan, L. (2012). Motivating English learners by helping them visualise their ideal L2 self: Lessons from two motivational programmes. *Innovation in Language Learning and Teaching*, *6*(2), 113–125. https://doi.org/10.1080/17501229.2011.614693

Moskovsky, C., Alrabai, F., Paolini, S., & Ratcheva, S. (2013). The effects of teachers' motivational strategies on learners' motivation: A controlled investigation of second language acquisition. *Language Learning*, *63*(1), 34–62. https://doi.org/10.1111/j.1467-9922.2012.00717.x

Papi, M., & Abdollahzadeh, E. (2012). Teacher motivational practice, student motivation, and possible L2 selves: An examination in the Iranian EFL context. *Language Learning*, *62*(2), 571–594. https://doi.org/10.1111/j.1467-9922.2011.00632.x

Safdari, S. (2021). Operationalizing L2 motivational self system: Improving EFL learners' motivation through a vision enhancement program. *Language Teaching Research, 25*(2), 282–305. https://doi.org/10.1177/1362168819846597

Sato, M. (2021). Generating a roadmap for possible selves via a vision intervention: Alignment of second language motivation and classroom behavior. *TESOL Quarterly, 55*(2), 427–457. https://doi.org/10.1002/tesq.611

Sato, M., & Lara, P. (2019). Interaction vision intervention to increase second language motivation: A classroom study. In M. Sato & S. Loewen (Eds.), *Evidence-based second language pedagogy* (pp. 287–313). Routledge. https://doi.org/10.4324/9781351190558-13

Sugita, M., & Takeuchi, O. (2010). What can teachers do to motivate their students? A classroom research on motivational strategy use in the Japanese EFL context. *Innovation in Language Learning and Teaching*, *4*(1), 21–35. https://doi.org/10.1080/17501220802450470

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Allyn & Bacon.

Thorsen, C., Henry, A., & Cliffordson, C. (2020). The case of a missing person? The current L2 self and the L2 motivational self system. *International Journal of Bilingual Education and Bilingualism*, *23*(5), 584–600. https://doi.org/10.1080/13670050.2017.1388356

Ushioda, E. (2016). Language learning motivation through a small lens: A research agenda. *Language Teaching, 49*(4), 564–577. https://doi.org/10.1017/S0261444816000173

Waninge, F., Dörnyei, Z., & De Bot, K. (2014). Motivational dynamics in language learning: Change, stability, and context. *The Modern Language Journal*, *98*(3), 704–723. https://doi.org/10.1111/modl.12118

Yashima, T. (2002). Willingness to communicate in a second language: The Japanese EFL context. *The Modern Language Journal*, *86*(1), 54–66. https://doi.org/10.1111/1540-4781.00136

Yashima, T., Nishida, R., & Mizumoto, A. (2017). Influence of learner beliefs and gender on the motivating power of L2 selves. *The Modern Language Journal*, *101*(4), 691–711. https://doi.org/10.1111/modl.12430

You, C., Dörnyei, Z., & Csizér, K. (2016). Motivation, vision, and gender: A survey of learners of English in China. *Language Learning*, *66*(1), 94–123. https://doi.org/10.1111/lang.12140

Research Forum

Significance Testing, Research Quality, and Second Language Research: A Reflection and Review

Imogen Custance
Osaka Jogakuin College and University

In any article reporting on quantitative research, you are likely to find the letter *p*. This letter, or rather what follows it, can draw the eye as a busy researcher seeks to decide whether the results presented are of use. Yet the desire to use this short-cut belies a history of problems. Though the field of second language research has made progress in moving away from this all-or-nothing, significant-or-not fixation, improving awareness of issues with statistical techniques is necessary. This article reviews some issues with significance testing to raise or reignite awareness in this commonly used statistic.

定量調査を報告する論文では、「p」という文字を頻繁に目にする。この文字の後に続く数字は、多忙な研究者が提示された研究結果が有用かどうかを判断しようとする際に、特に注目される。しかし、このショートカットを使いたいという願望には、問題の歴史が潜んでいる。第二言語研究の分野では、このようなオール・オア・ナッシング、有意か無か、といった2元的な固定観念から脱却しつつあるが、統計的手法に関する問題意識を向上させることは必要である。本稿では、有意性検定に関するいくつかの問題を検討し、この広く使用されている統計量に対する認識を高める、あるいは喚起させることを目的とする。

Keywords: methodology; research methods; research quality; state of the field

https://doi.org/10.37546/JALTJJ47.1-4

JALT Journal, Vol. 47, No. 1, May 2025

In 2016, the American Statistical Association took the unprecedented step of publishing a statement on the use of *p* values in research (Wasserstein & Lazar, 2016). This step was taken in response to "highly visible discussions" (p. 129) regarding the use of null hypothesis significance testing (NHST) in a wide range of fields and to "draw renewed and vigorous attention to changing the practice of science with regards to the use of statistical inference" (p. 130). Yet as part of the statement they also emphasised that nothing mentioned was new information and that "statisticians and others have been sounding the alarm" about issues related to significance testing "for decades, to little avail" (p. 130). Indeed, over fifty years ago Bakan (1966) was already stating that the arguments in his paper were "hardly original" (p. 423) and in the following decade Carver suggested "educational research would be better off if it stopped testing its results for statistical significance" (1978, p. 378).

The aim of this article is not to call for an end to the use of statistical significance testing within second language research. Rather, it is to critically evaluate its use within the field and demonstrate the positive impact that greater consideration of issues associated with significance testing has had in what is hoped to be an accessible way. To this end, background information about the nature of statistical significance testing and some of the issues associated with it are introduced first. This is followed by a more detailed examination of its use in second language research and the impact a reliance on statistical significance testing as the main method for conducting quantitative analysis has had on the field. Next is an acknowledgement of the efforts being made to address these issues and the resulting changes in research quality, but also a recognition that increasing use of more advanced statistical techniques necessitates careful consideration of how research is used and by whom. I conclude that a shift in mindset regarding statistical significance testing is more appropriate than calling for the cessation of its use.

Background

Statistics has an influence on almost every aspect of modern life (Hand, 2008). It is used to make decisions and predictions about the future; to try and elucidate the relationships that underlie our reality. At its core, statistics is about modelling the world around us in such a way that we can understand it better. A model can never, however, be a perfect representation of reality. As such, there must be an inherent acceptance of uncertainty in any statistical model and in the results of a statistical test. Understand-

ing the impact that this uncertainty has on the interpretation of results is necessary for those who wish to use statistics appropriately.

Null hypothesis significance testing (NHST) is one way in which uncertainty is acknowledged. In this type of testing, an assumption is made and termed the *null hypothesis*, H_0. This assumption is usually taken to be that there is no difference between two groups, with the actual idea of interest being that something has, in fact, caused there to be a difference between them. NHST is then used to check if the data that have been observed would be consistent if the null hypothesis were true. If the probability of the data occurring given the null hypothesis is true is below a predetermined alpha level α, usually α =.05 or α = .01, the null hypothesis is rejected. The probability is termed the *p* value, and "describes the probability that we would observe the value of the test statistic as extreme or more extreme than that actually observed, if the null hypothesis were true" (Hand, 2008, p. 89).

When calculating statistical significance using NHST, researchers must be careful to acknowledge the two possible mistakes that could arise when examining the *p* value and deciding whether to reject the null hypothesis. The first, a Type I error, is rejecting the null hypothesis when it is in fact true. The second, a Type II error, is failing to reject the null hypothesis when an alternate hypothesis is true. Hand (2008) explained these two types of error using the example of a court of law in which the null hypothesis is the assumption of innocence, with a Type I error the equivalent of an innocent person being found guilty, and a Type II error the equivalent of someone who is actually guilty being declared innocent. Norris (2015) highlighted that NHST, because only the null hypothesis is considered, will always help to avoid a Type I error—with a small enough alpha level, we can be relatively certain that the null hypothesis should be rejected. However, reducing the chance of a Type II error requires careful consideration of the statistical power necessary for a study. Field (2017) defines power as the "the ability of a test to find an effect" when there is an effect to be found and depends on the size of the effect, the sample size, and alpha-level or corrected alpha-level if multiple tests are conducted (p. 84).

While accepting that the *p* value can be useful because it helps researchers "be cautious in claiming that a difference or relationship they have observed in their sample data is actually rare…in comparison with the assumption that there is no such pattern" (Norris, 2015, p. 100), Norris is very critical of its use, in part because it is often misinterpreted as doing much more than this. Carver (1978) presented three "fantasies" (p. 383)

about what information *p* values provide. These are misinterpreting the *p* value as the probability that the results were due to chance, the probability that a replication of the study would achieve the same result, and the probability that the research hypothesis is true. Carver, along with others (e.g., Field, 2017; MacInnes, 2022), also emphasised the importance of understanding that the probability of the null hypothesis being true given the data, $p(H_0|D)$, cannot be inferred from the probability of the result of the NHST, which is the probability of the data given the null hypothesis is true, $p(D|H_0)$. Other misinterpretations about *p* values raised by Greenland et al. (2016) include their ability to demonstrate whether a particular hypothesis is true or false, that the size of the *p* value itself is indicative of how strong the evidence for/against the null hypothesis is, and that finding statistical significance indicates an important discovery or observation has been made.

The reporting of effect sizes, which are standardized measures of how large or seemingly important a difference that has been identified is, has been advocated as a way to move away from a focus on NHST. Effect sizes not only give a measure of the importance of a discovery, but are also unaffected by sample size (Field, 2017). This is important because *p* values depend on the size of the sample (Field, 2017; Norris, 2015; Plonsky, 2015) and given a large enough sample size, it will always be possible to find a statistically significant difference between populations (Bakan, 1966). This means that whether a result is considered significant or not can be a function of the sample size, rather than the existence of an actual, meaningful difference or effect between groups. An effect size reported with a confidence interval provides more meaningful information about a result than a *p* value (Field, 2017; Norris & Ortega, 2000; Plonsky, 2015).

Misinterpretation and/or a lack of understanding of *p* values is an issue, but one that could potentially be solved through increased education. Indeed, the purpose of Greenland et al.'s (2016) paper was to provide a resource to help researchers "avoid and spot misinterpretations" (p. 337). However, researchers in a range of fields have gone beyond calling for increased understanding of *p* values and instead suggested that the use of NHST should be actively discouraged or stopped, with some journals banning its use (Trafimow & Marks, 2015). In the following section, I present some of the reasons given for such proposals, with reference to the use of NHST in second language research.

Issues With NHST in Second Language Research

It has been argued (e.g., Labaree, 2011; Porter, 1996) that professions with more applied and practical purposes "find themselves subject to the greatest external pressures and the strongest need to demonstrate the credibility of their claims through quantitative means" (Labaree, 2010, p. 624). The use of experimental design and quantitative analysis of the results is strongly associated with the scientific method, which in turn carries positive connotations of objectivity and trustworthiness. This might in part explain the prevalence of NHST in a range of fields, including second language research, despite the limited nature of the information provided by such tests. The use of quantitative research methods and an apparently objective method of determining the significance of claims can be seen to help legitimize a field which can in turn promote investment and development. Norris (2015) suggested that "The simplicity and apparent certainty of significance testing is alluring" (p. 101), which has made it a popular way to obtain and interpret quantitative research results. Field (2017) similarly suggested that "NHST seems to provide an easy way to disentangle the 'correct' conclusion from the 'incorrect' one" and that it is "appealing to teach" because students "can follow the rule that a $p < 0.05$ is 'significant' and a $p > 0.05$ is not" even if they do not understand the underlying logic of the test (p. 97).

Yet it has been suggested that the emphasis placed on achieving statistical significance in research results has in fact moved fields such as educational research (Carver, 1978), psychology (Chambers, 2017), and second language research (Plonsky, 2015) away from the ideals of the scientific method and scientific rigor. One of Carver's (1978) central arguments for stopping testing results for statistical significance is that it has led to an overemphasis on the finding of statistical, as opposed to scientific, significance in results. This is an argument echoed by Plonsky (2013, 2015), who further suggested that the tendency to not report non-significant results belies an underlying fixation on achieving statistical significance rather than examining what can be understood from the data collected.

A fixation on achieving statistical significance is problematic for a variety of reasons. First, the commonly used cut-off for having obtained a "significant" result, $p < 0.05$, is arbitrary (Plonsky, 2015) and "encourages all-or-nothing thinking" (Field, 2017, p. 99). Rather than considering the actual size of an effect, NHST encourages a knee-jerk reaction as to whether the data should be examined in more detail or not, or even reported at all. Second, it is likely that there is a bias towards publishing

results that are found to be statistically significant (Marsden et al., 2018). According to Norris (2015), there "seems to be an artificial imbalance" (p. 104) between the number of studies finding statistical significance and those failing to within L2 research. Whether this is due to researchers deciding not to submit a study where no statistical significance was found, or journal editors rejecting it for perceived lack of importance, these results are "put away" leading to a skewed understanding of the research domain (Norris & Ortega, 2006, p. 21). Finally, the apparent preference for publishing positive, exciting results might lead researchers to engage in questionable research practices in order to have their work published (Chambers, 2017; Field, 2017; Marsden et al., 2018). This includes practices such as *p*-hacking, where researchers make decisions regarding what data or analyses to use, and which results to report, based on whether they yield statistically significant results, and HARKing (hypothesizing after the results are known), where a hypothesis is presented as having been made before data was collected or analysed when this was not the case (Field, 2017).

The aim of research is to understand and develop theories that explain the observations that are made. Research results can provide evidence for new theories; they can support or contradict previous findings. However, the use of NHST as a gatekeeper to publication is "a corrupt form of the scientific method" (Carver, 1978, p. 378). When publication is determined, or thought to be determined, by whether or not statistical significance is found, the process of theory falsification that is central to the scientific process falls apart. A negative result is considered "trash" (Plonsky, 2015, p. 24), is less likely to be published, and will not have an impact on the development of theory or future research. This can lead to the existence of *undead theories* (Ferguson & Heene, 2012), theories that continue to be used because negative evidence against them remains unpublished, or that otherwise resist attempts at falsification. An unexpected result of no statistical significance does not mean a study is necessarily uninformative or unimportant, or that it should not be considered for publication.

The presence of publication bias and an apparent disregard for results of non-significance started to receive more attention as L2 researchers began to look at conducting meta-analyses of primary research (e.g., Norris & Ortega, 2000). Meta-analyses are a powerful means of understanding the findings in a field of research. Whereas the findings of individual studies might be "attributable to chance variability as well as idiosyncrasies in design, analysis, sampling error [and] research setting" (Norris & Ortega,

2000, p. 423), a secondary analysis combining the results of several primary studies can help bring to light overall patterns that are applicable beyond the setting of any individual study. Publication bias is problematic for meta-analyses for the same reasons it is problematic to the field as a whole—if non-significant results are not published, they cannot be included in meta-analyses, and any patterns discerned in the meta-analyses will not reflect the reality that some studies have not found significance. There is also an issue when statistics, especially standard deviations or effect sizes, are not reported for non-significant results, a tendency identified by Plonsky (2013). The way in which meta-analyses are generally conducted is by finding primary data related to a treatment or condition and estimating the overall magnitude of any observed relationship or effect across the different studies. This involves an examination of the effect sizes in the primary studies. However, if the effect sizes or data required for calculating effect sizes, i.e., standard deviations, are not reported, it is not possible to include a study in a meta-analysis. Thus, a failure to report non-significant results or report all relevant statistics creates a significant barrier to the undertaking of meta-analytic research and the furtherance of the field through secondary research. Many of the issues with NHST mentioned so far are also underscored by Norris and Ortega (2000).

Reporting of results and questionable research practices are not the only issues associated with NHST in the field. Plonsky (2013, 2014) highlighted issues with the appropriacy of some of the statistical tests used in L2 research. He reported that the most commonly used inferential statistical test within the field from 1990–2010 was ANOVA, with 56% of studies using this type of analysis. Plonsky (2013, 2014) suggested that using a means-based test like ANOVA is problematic because they can obscure some of the information that is of interest to the field. When conducting an ANOVA, the means for the different groups in the analysis are taken and compared. By taking the mean, information relating to variance between those within the group is necessarily lost. Plonsky argued that given the complex nature of the constructs involved in L2 research, failing to preserve this variance reduces the conceptual validity of results obtained with these types of tests—comparing means is a practice that "sacrifices variance, informational richness, and statistical power for an analytic model that appears more straightforward" (Plonsky, 2014, p. 453).

A further issue with means-based testing is that researchers often conduct multiple tests which has a "debilitating effect on statistical power" (Plonsky, 2014, p. 453) as the alpha value must be adjusted for each

additional test conducted. Indeed, low statistical power is thought to be a major concern for the field for a number of reasons (Lindstromberg, 2023; Plonsky, 2013). Statistical power is affected by both sample size and the expected effect size. As such, a study is likely to be underpowered if the sample sizes and/or the effect sizes is small. This is an issue with second language research because sample sizes are typically small (Lindstromberg, 2016; Plonsky, 2013). When a study is underpowered, the risk of a Type II error, failing to find a significant effect when there is one, is increased. Plonsky (2015) suggested that a power level of 0.8, or an 80% chance of detecting a real effect, is appropriate for social science research. However, in his study of research quality, Plonsky (2013) estimated that statistical power in the field, based on the median sample and effect sizes of the publications examined, was just 57% on average. In addition, research in the related fields of psychology and education have suggested that the effect sizes found in second language research are likely to be overestimated as a result of publication bias (Lindstromberg, 2023). As such, researchers might require even larger sample sizes to achieve sufficient statistical power when using NHST.

In sum, the number of issues related to statistical significance testing in second language research is indicative that it "is probably not well-conceived or accurately interpreted" (Norris, 2015, p. 106) and has resulted in a wide range of problems for the field.

The Impact of Publication Requirements

Despite the issues mentioned in the previous section, NHST continues to be used. However, changes in editorial policies and publication guidelines from top journals have started to deemphasise NHST in favour of procedures that highlight the scientific significance of results and academic rigour. The most recent American Psychological Association (APA) guidelines (American Psychological Association, 2019) contain a chapter focused on journal article reporting standards that details what information should be included for different types of research. These standards, initially developed in 2008 for quantitative research (APA Publications and Communications Board Working Group on Journal Article Reporting Standards, 2008) and revised in 2018 (Appelbaum et al., 2018), include the reporting of intended sample size and the statistical power analysis used to determine it, descriptive statistics, effect sizes, and exact *p*-values for all statistical tests whether a significant effect has been found or not. When establishing the initial standards in 2008, the working group

aimed to create guidelines that would "[promote] sufficient and transparent descriptions of how a study was conducted and what researcher(s) found [...to permit] the users of the evidence to judge more accurately the appropriate inferences and applications derivable from research findings" (APA Publications and Communications Board Working Group on Journal Article Reporting Standards, 2008, p. 847). They also highlighted that the suggested standards could encourage researchers to consider study plans more carefully, facilitate replication studies, and increase the number of studies that can be included in meta-analyses. The development of explicit reporting standards based on these ideas, in addition to calls for increased use of open science practices (e.g., Al-Hoorie et al., 2024; Liu et al., 2023; Marsden & Morgan-Short, 2023), demonstrate a desire to acknowledge and overcome some of the issues listed in the previous section.

One way to see the extent to which changes to publishing standards have impacted the field is to revisit articles published before they were introduced and consider how the research might be done differently if conducted now. A journal in second language research at the forefront of some guideline changes is *Language Learning,* which has required the reporting of effect sizes since 2000 (Ellis, 2000), specified a range of guidelines for reporting quantitative research in 2015 (Norris et al., 2015), and introduced registered reports, whereby a study can essentially be approved for publication before results are known in 2018 (Marsden et al., 2018). In 1999, the editor's statement highlighted that one of the journal's strengths was the "high quality of its empirical research" and "focus on the systematic collection, analysis and evaluation of data" (Ellis, 1999, p. vi). In the same volume, an article, Skehan and Foster (1999), reporting the results of an experiment examining the effect of task structure and processing load, was published.

In their study, Skehan and Foster used a 2 x 4 between-subjects design to investigate two tasks and four performance conditions. The nature of the tasks and measures used, including specific reasons for why certain measures were not used are explained, and descriptive statistics were reported along with group sample sizes. Participants were randomly assigned to task and condition, and measures were checked for collinearity. There is probably sufficient detail included within the article to make replication of the study possible, as required by the editor (Ellis, 1999).

However, only 47 participants were involved in the study and whether the study had enough statistical power for the two-way ANOVAs conducted should be considered. Using the Plonsky and Oswald (2014) suggested

benchmarks for interpreting effect sizes (Cohen's *d*) for L2 research of small (*d* = 0.4), medium (*d* = 0.7), and large (*d* = 1.0) it is possible to calculate the sample size necessary to achieve effect sizes of these levels. I used the computer program G*Power (Faul et al., 2009). The alpha level was set to .05, and power to .8, the level suggested for social science research (Plonsky, 2015). G*Power calculates the *N* size based on a Cohen's *f*, so I converted the *d* values suggested by Plonsky and Oswald (2014) using the formula

$$f^2 = \frac{d^2}{2k}$$

(Statistics How To, n.d.) where *k* is the number of groups (*k* = 8 for Skehan and Foster's study). Table 1 shows the effect sizes (*f* and *d* values), and the *N* sizes suggested to be necessary if these effect sizes were expected. They indicate that even if a large effect size was expected, the study would have been underpowered given the actual sample size (*N* = 47).

Table 1

Estimation of Necessary N-Sizes for Skehan and Foster's (1999) Study

Effect Size	Cohen's *d*	Cohen's *f*	*N* size
Small	0.4	0.100	1095
Medium	0.7	0.175	360
Large	1.0	0.250	179

An early consideration of statistical power within Skehan and Foster's study design process might have helped the researchers to adjust their research design. The researchers could have examined fewer performance conditions, a strategy recommended by Norris and Ortega (2000), or chosen to analyze the data differently. Alternatively, they might have decided to conduct additional data collection to achieve a more appropriate *N* size. In either case, the study would likely have been improved, highlighting why addressing intended sample size is beneficial to research.

There are also differences in what and how the statistical results might be reported today. First, no information was given about the results of assumption checks on the data. Though the researchers might have conducted these, by not reporting the results, readers cannot judge whether the *p*-values presented are accurate. At the time the article was published,

ANOVA was considered a robust test, so issues with assumption violations might not have been considered, but this would not be the case now. Second, even if the data met the assumptions required to produce accurate *p*-values, no effect sizes were reported, nor any focus given to how meaningful the statistically significant differences between tasks and performance conditions were. The arguments made in the discussion would have been strengthened were they supported by effect size information. Finally, visualizations of the data, including effect sizes and confidence intervals, would potentially have made interpretation of the results simpler.

Where Do We Stand?

To an extent, by not stopping testing results for statistical significance when the issues with NHST were first raised, "a great deal of mischief" (Bakan, 1966, p. 423) and damage has probably been wrought in various fields, including that of second language research. NHST has become the "go-to analytic approach…for making sense of numerical data" (Norris, 2015, p. 97) as a result of a self-fulfilling cycle whereby it is "[taught] because it's what we do; …[done] because it's what we teach" (Wasserstein & Lazar, 2016, p. 129). The misuse and misinterpretation of *p* values over the years has likely resulted in a range of somewhat erroneous theories gaining traction while the failure to report non-significance or details related to non-significant results has harmed the field's ability to conduct meta-analyses. However, there are clear indications that the issues raised all those years ago are now being addressed much more proactively, as exemplified by the various changes to editorial policies within second language research journals, the publication of books focused on the use of statistics within L2 research, the increase in research that has been conducted into study quality, and the launch of a journal, *Research Methods in Applied Linguistics*, that is "devoted exclusively to the study and advancement of methods and approaches in language-related research" (Li & Prior, 2022, p. 1). The shift from a focus on a significant result to ensuring more transparent reporting practices and how a finding might impact our understanding of theory is positive.

However, the full impact of these calls for change must not be underestimated. The field of second language research is maturing, as partly demonstrated by the increased use of more advanced statistical techniques. Khany and Tazik (2019) found that the number of research articles requiring the knowledge of intermediate or advanced techniques increased from 20.61% between 1986–1995 to 39.08% between 2006–2015. While this is far

from a negative development, it necessitates an examination of how this research is used and by whom. Loewen et al. (2020) found "mixed evidence of…researchers' ability to use and interpret" (p. 883) statistical information and "limited overlap" (p. 884) between self-perceived statistical knowledge and actual knowledge as determined by a test. They suggested this is likely to make calls for increased rigour and use of more advanced techniques difficult to achieve. If the use of more advanced techniques is being advocated, it is necessary to improve statistical literacy which is "critical for both producers and consumers of L2 research" (Gonulal et al., 2017, p. 4). While courses that cover statistical techniques can be effective in raising learners' confidence in their ability to interpret and use such techniques (Gonulal et al., 2017), whether institutions are able to offer such courses, or professionals take them, will impact how well the field can adapt to the presence of increasingly complex statistics.

Most of the arguments from second language researchers presented in this paper are from those aligned with the field of second language acquisition. Gass et al. (2021) argued that the area of second language acquisition has "evolved into a unique area of inquiry seeking to understand how second languages are learned without an emphasis on how they are taught" (p. 247). However, the knowledge and theory produced in this field continues to have an impact on that of language teaching as "the two disciplines feed one another and are relevant to one another" (Gass et al., 2021, p. 247). Not considering how to maintain the accessibility of this research to members of the overarching field of second language research has the potential to cause a different type of damage to that resulting from misuse of *p* values. The use of *p* values can make the results of a study more accessible to other L2 researchers. Effectively, by including *p* values, the reporting of results remains in line with what most researchers in the field would expect and find easy to interpret. In this way, continued reporting of *p* values makes it possible for more people to learn something from a publication. While there might be some misinterpretation of what the *p* value means, by keeping this avenue of interpretation open, the field itself remains more open and inclusive.

Of course, there are other ways in which knowledge and expertise develop. Just as Isaac Newton saw further by standing on the shoulders of giants, so does anyone in a field learn from what has gone, or been published, before. This is not limited to the results of a study, but also the methods and analyses employed to answer the research questions posed. Thus, greater attention to, and reporting of, whether assumptions are

met before conducting a statistical test highlights the importance of this issue to researchers. The shift in focus towards not just the results, but the appropriacy of methods used is already evident within the field. The introduction of registered reports (e.g., Marsden et al., 2018) is another step in the direction of emphasizing study quality and confidence in results over how novel the results observed are. Yet these facts do not preclude the use of statistical significance testing. Rather, they emphasize that it is necessary to consider whether a given test is appropriate for the data obtained or relevant for answering the research questions posed.

Conclusion

In conclusion, and as many have said before (e.g., Bakan, 1966; Norris, 2015; Wasserstein & Lazar, 2016), statistical significance testing itself is not at fault; the problem lies with how it has been used and overemphasised. With this in mind, it is necessary that journals and journal reviewers take special care to check that publications do not risk propagating false ideas regarding these tests, and that researchers and graduates are educated to ensure that they understand the limitations of statistical significance. Brown's (2016) *Statistics Corner* is a collection of articles published in the JALT Testing and Evaluation SIG's publication that is a useful resource in this regard. In addition, when planning a study, researchers should carefully consider the statistical power necessary for conducting specific tests, and report descriptive statistics including means and standard deviations, effect sizes, and exact *p*-values for all results, including those that are non-significant. The field and research techniques used within it will continue to evolve and time will tell if NHST "survives as a useful technique or is replaced" (Norris, 2015, p. 123). For now, it continues to play a role and perhaps needs to as a simple, if somewhat limited, part of the research landscape.

Imogen Custance is a lecturer at Osaka Jogakuin University and College. Her research interests include the development of speaking skills and communicative competence.

References

Al-Hoorie, A. H., Cinaglia, C., Hiver, P., Huensch, A., Isbell, D. R., Leung, C., & Sudina, E. (2024). Open science: Considerations and issues for TESOL research. *TESOL Quarterly*, *58*(1), 537–556. https://doi.org/10.1002/tesq.3304

American Psychological Association. (2019). *Publication manual of the American Psychological Association* (7th ed.).

APA Publications and Communications Board Working Group on Journal Article Reporting Standards. (2008). Reporting standards for research in psychology: Why do we need them? What might they be? *American Psychologist*, *63*(9), 839–851. https://doi.org/10.1037/0003-066X.63.9.839

Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *American Psychologist*, *73*(1), 3–25. https://doi.org/10.1037/amp0000191

Bakan, D. (1966). The test of significance in psychological research. *Psychological Bulletin*, *66*(6), 423–437. https://doi.org/10.1037/h0020412

Brown, J. D. (2016). *Statistics corner: Questions and answers about language testing statistics*. TEVAL SIG.

Carver, R. (1978). The case against statistical significance testing. *Harvard Educational Review*, *48*(3), 378–399. https://doi.org/10.17763/haer.48.3.t490261645281841

Chambers, C. (2017). *The seven deadly sins of psychology: A manifesto for reforming the culture of scientific practice.* Princeton University Press. https://doi.org/10.1515/9781400884940

Ellis, N. C. (1999). Editor's statement. *Language Learning*, *49*(1), v–vi. https://doi.org/10.1111/1467-9922.00068

Ellis, N. C. (2000). Editor's statement. *Language Learning*, *50*(3), xi–xiii. https://doi.org/10.1111/0023-8333.00135

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Ferguson, C. J., & Heene, M. (2012). A vast graveyard of undead theories: Publication bias and psychological science's aversion to the null. *Perspectives on Psychological Science*, *7*(6), 555–561. https://doi.org/10.1177/1745691612459059

Field, A. (2017). *Discovering statistics using IBM SPSS statistics* (5th ed.). Sage.

Gass, S., Loewen, S., & Plonsky, L. (2021). Coming of age: The past, present, and future of quantitative SLA research. *Language Teaching*, *54*(2), 245–258. https://doi.org/10.1017/S0261444819000430

Gonulal, T., Loewen, S., & Plonsky, L. (2017). The development of statistical literacy in applied linguistics graduate students. *ITL: International Journal of Applied Linguistics*, *168*(1), 4–32. https://doi.org/10.1075/itl.168.1.01gon

Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., & Altman, D. G. (2016). Statistical tests, *p*-values, confidence intervals, and power: A guide to misinterpretations. *European Journal of Epidemiology*, *31*(4), 337–350. https://doi.org/10.1007/s10654-016-0149-3

Hand, D. J. (2008). *Statistics: A very short introduction*. OUP Oxford. https://doi.org/10.1093/actrade/9780199233564.001.0001

Khany, R., & Tazik, K. (2019). Levels of statistical use in applied linguistics research articles: From 1986 to 2015. *Journal of Quantitative Linguistics*, *26*(1), 48–65. https://doi.org/10.1080/09296174.2017.1421498

Labaree, D. (2011). The lure of statistics for educational researchers. *Educational Theory*, *61*(6), 621–632. https://doi.org/10.1007/978-90-481-9873-3_2

Li, S., & Prior, M. T. (2022). Research methods in applied linguistics: A methodological imperative. *Research Methods in Applied Linguistics*, *1*(1), 100008. https://doi.org/10.1016/j.rmal.2022.100008

Lindstromberg, S. (2016). Inferential statistics in Language Teaching Research: A review and ways forward. *Language Teaching Research*, *20*(6), 741–768. https://doi.org/10.1177/1362168816649979

Lindstromberg, S. (2023). The winner's curse and related perils of low statistical power – spelled out and illustrated. *Research Methods in Applied Linguistics*, *2*(3), 100059. https://doi.org/10.1016/j.rmal.2023.100059

Liu, M., Chong, S. W., Marsden, E., McManus, K., Morgan-Short, K., Al-Hoorie, A. H., Plonsky, L., Bolibaugh, C., Hiver, P., Winke, P., Huensch, A., & Hui, B. (2023). Open scholarship in applied linguistics: What, why, and how. *Language Teaching*, *56*(3), 432–437. https://doi.org/10.1017/S0261444822000349

Loewen, S., Gönülal, T., Isbell, D. R., Ballard, L., Crowther, D., Lim, J., Maloney, J., & Tigchelaar, M. (2020). How knowledgeable are applied linguistics and SLA researches about basic statistics?: Data from North American and Europe. *Studies in Second Language Acquisition*, *42*(4), 871–890. https://doi.org/10.1017/S0272263119000548

MacInnes, J. (2022). *Statistical inference and probability*. Sage. https://doi.org/10.4135/9781529682748

Marsden, E., & Morgan-Short, K. (2023). (Why) Are open research practices the future for the study of language learning? *Language Learning, 73*(2), 344–387. https://doi.org/10.1111/lang.12568

Marsden, E., Morgan-Short, K., Trofimovich, P., & Ellis, N. C. (2018). Introducing registered reports at Language Learning: Promoting transparency, replication, and a synthetic ethic in the language sciences. *Language Learning*, *68*(2), 309–320. https://doi.org/10.1111/lang.12284

Norris, J. M. (2015). Statistical significance testing in second language research: Basic problems and suggestions for reform. *Language Learning*, *65*(S1), 97–126. https://doi.org/10.1111/lang.12114

Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, *50*(3), 417–528. https://doi.org/10.1111/0023-8333.00136

Norris, J. M., & Ortega, L. (2006). The value and practice of research synthesis for language learning and teaching. In J. M. Norris & L. Ortega (Eds.), *Synthesizing research on language learning and teaching* (pp. 1–50). John Benjamins. https://doi.org/10.1075/lllt.13.04nor

Norris, J. M., Plonsky, L., Ross, S. J., & Schoonen, R. (2015). Guidelines for reporting quantitative methods and results in primary research. *Language Learning*, *65*(2), 470–476. https://doi.org/10.1111/lang.12104

Plonsky, L. (2013). Study quality in SLA: An assessment of designs, analyses, and reporting practices in quantitative L2 research. *Studies in Second Language Acquisition*, *35*(4), 655–687. https://doi.org/10.1017/S0272263113000399

Plonsky, L. (2014). Study quality in quantitative L2 research (1990–2010): A methodological synthesis and call for reform. *The Modern Language Journal*, *98*(1), 450–470. https://doi.org/10.1111/j.1540-4781.2014.12058.x

Plonsky, L. (2015). Statistical power, *p* values, descriptive statistics, and effect sizes. In L. Plonsky (Ed.), *Advancing quantitative methods in second language research* (1st ed., pp. 23–45). Routledge. https://doi.org/10.4324/9781315870908-3

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning*, *64*(4), 878–912. https://doi.org/10.1111/lang.12079

Porter, T. M. (1996). *Trust in numbers: The pursuit of objectivity in science and public life*. Princeton University Press. https://doi.org/10.1515/9780691210544

Skehan, P., & Foster, P. (1999). The influence of task structure and processing conditions on narrative retellings. *Language Learning*, *49*(1), 93–120. https://doi.org/10.1111/1467-9922.00071

Statistics How To. (n.d.). *Cohen's f statistic: Definition, formulas*. Retrieved 10 August 2024, from https://www.statisticshowto.com/cohens-f-statistic-definition-formulas/

Trafimow, D., & Marks, M. (2015). Editorial. *Basic and Applied Social Psychology*, *37*(1), 1–2. https://doi.org/10.1080/01973533.2015.1012991

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA statement on *p*-values: Context, process, and purpose. *The American Statistician*, *70*(2), 129–133. https://doi.org/10.1080/00031305.2016.1154108

Reviews

What English Language Teachers Need to Know Volume III: Designing Curriculum (2nd ed). MaryAnn Christison and Denise E. Murray. Routledge, 2022. xiv + 376 pp. ¥8,200. https://doi.org/10.4324/9780429275746

Reviewed by
Kenneth Charles Lambo
Carleton University

The second edition of MaryAnn Christison and Denise E. Murray's (2022) "What English Language Teachers Need to Know Volume III: Designing Curriculum" is a comprehensive resource that delves into the question of "What do teachers need to know and be able to do in order to help their students to learn English?" (p. i). For this edition, the authors have expanded their chapters to address the pressing technological and multilingual challenges facing learners and teachers in today's English Language Teaching (ELT) education, making it a timely and relevant reference for information. It is a must-read for pre-service teachers, policymakers, and graduate students interested in language education across different contexts. It takes us through the theory and practice in ELT curricula, aiming to develop, design, and promote student learning as the main goal of the curriculum.

The book's emphasis on the intricate status of English across various countries is noteworthy, as it takes into account the cultural, political, and historical contexts in which English is utilized. This consideration enables the presentation of multiple teaching methodologies tailored to the diverse contexts and needs of language learners. For example, the pedagogical approach in countries where English serves as a second language may differ significantly from that in countries where it is taught as a foreign language. The book is organized around three pillars of teaching—planning, instruct-

ing, and assessing. This three-pillar model is not linear but reiterative, with each pillar constantly evolving and interacting with each other (Graves, 2008; Macalister & Nation, 2011; 2020). The centre of this tripartite process is learning, which puts the content, teachers, and learners at the heart of the dynamic process of iteration, reevaluation, and curricular innovation.

In Part I, the authors show how various perspectives on education, language, and learning, along with input from different stakeholders, shape the curriculum and influence what gets included in it. This part contains five chapters to set the stage on the various ways a curriculum is developed (Chapter 1), the sociocultural milieu (Chapter 2), the decision-making behind the choice and use of a certain approach (Chapter 3), how this decision translates to teaching in a multilingual group of learners (Chapter 4), and how learning through the aid of technology is an important consideration in the 21st-century curriculum (Chapter 5).

Part II of the book contains a practical guide that discusses the curriculum design process, showing how to design a curriculum tailored to specific situations. In this part, chapters explain the cycle of curriculum design (Chapter 6), the connection of the curriculum to the course/program (Chapter 7), and the evaluation scheme on the quality of the curriculum (Chapter 8). The remaining parts—III, IV, V, and VI—offer examples of various approaches to curriculum choices. They focus on language, content, the learner, and the learning process. These sections are grounded in current research in ELT and related fields. Starting in Part III, this section focuses exclusively on linguistic-based curricula. It comprises six approaches (i.e., structural, notional-functional, genre/text-based, academic language functions, vocabulary, and language skills approach) based on certain language features, such as grammar. This practical guide equips teachers with toolkits to design effective sequencing of grammatical structures within a communicative framework, ensuring they are well-prepared for their teaching roles.

Part IV focuses on content-based curricula. This section departs from the focus on language and is centrally developed for *content*. In this part, there are claims that language and content learning complement each other, where the more content is learned the more learners improve their language. It is divided into two main approaches: content and language integrated (Chapter 15) and topical and situational (Chapter 16). The former incorporates essential content, such as academic subjects in K–12 schools or at the tertiary level, and the latter selects content based on what is motivating and valuable for learners. Each type offers a range of imple-

mentations, and the distinction between them is not always clear-cut. What stands out most about content-based curricula is that language use, in all its complexity, is guided by the linguistic demands of the content.

In Part V, the authors explain the concept of learner-centred curricula, which prioritize the learning *process* over the content. This means that the curriculum is designed around how learners learn rather than the specific goals of what they should learn. This book section includes three chapters focusing on negotiated, humanistic, and task-based curricula. Negotiated curricula enable the learners to be in the driver's seat, empowering them to be autonomous and putting the responsibility on the teachers for curriculum development because each class differs in collective needs and linguistic goals. The humanistic curricula have similarities with the previous, but the balance is different. The teachers are facilitators in the learning process, guiding students to discover knowledge independently. A task-based approach to curriculum design centres on tasks, recognizing that language is a tool learners use to engage with others and, in the process, use the language more naturally. This approach is particularly beneficial for language education as it allows students to learn daily task-related activities, promoting a deeper understanding and retention of the language.

In Part VI, the authors transition on the focus on process to a focus on *product*. This part of the book discusses three approaches under this curriculum type: outcome-based, competency-based, and standards-based. Each approach emphasizes outcomes, moving away from the foci of the previous curricula mentioned in the book. The outcome-based approach focuses on the desired results of the learning process, ensuring that students achieve specific learning outcomes. The competency-based approach emphasizes the skills and knowledge that learners should acquire, focusing on the development of specific competencies. The standards-based approach sets specific criteria for what learners should know and be able to do, ensuring that students meet certain proficiency standards. Understanding these approaches is important for educators as it helps them set clear learning objectives and assess the effectiveness of their teaching.

As elucidated in the book's preface, English language teaching within the global context encounters ongoing challenges that continuously influence curriculum development. In certain regions, English is valued for its substantial economic impact, often facilitated through shadow education and tutoring (Cao, 2024). Conversely, in regions classified as part of the inner circle—namely, Britain, Australasia, and North America (Kachru, 1986)—language curricula are increasingly adapting to serve a multilingual

clientele, the learners, coming from all linguistic backgrounds as a product of the increased global human mobility. The book acknowledges the global impact of English as it is "consumed and transformed transnationally" (p. ix). This global impact of English has significant implications for language education, as it necessitates a curriculum that is adaptable and inclusive, catering to the diverse linguistic backgrounds of learners. As a developing curriculum expert/researcher and an adult language educator, I find this a step forward to embracing and addressing the plurilingual and linguistic panorama of our classrooms (Gazzola et al., 2023; Piccardo, 2013).

On the other hand, Christison and Murray are well-known internationally for their contributions to the field of ELT. Their reputation is accompanied by a global field experience through their vignettes that are easy to read and relatable for novice and experienced teachers. The accessibility of the book resonates not just with me as a Canadian language educator but also with readers around the world who have different values and perspectives on what the best language education means to them. We have the option, not as passive readers of information, to apply a suitable approach. The intended readers—specifically, the language teachers—have the agency to choose, play around, evaluate, and examine the approaches until the "right" one works out fine. Teachers become the final frontier in curriculum success, acting as the "intermediary between national curriculum and classroom" (Parent, 2011, p. 186). According to the authors, examining the iterative curriculum development process underscores the importance of the educational experiences co-constructed by teachers and learners. This reciprocal relationship suggests that the effectiveness of the curriculum can be assessed when both teachers and learners are satisfied with the educational outcomes, which are central to the classroom environment—the "heart of education" (Graves, 2008, p. 152).

Despite the authors' effort to be comprehensive, novice teachers may find the contents needing an in-depth explanation. Although the book is presented as being of the "how-to" genre, there may be foundational knowledge missing. There are two examples that I find needing more contextualization. First, readers may ask about curricular innovation because they work in an institution with an outdated curriculum. Chapter 6 briefly discusses curricular innovation, which may not be enough to support novice teachers looking into the complex process of educational change. Second, the curriculum is not a one-person job; stakeholders (e.g., parents, policymakers, and teachers) have interests to escalate within the commonplaces of the curriculum—"subject matter, milieu, learner, and

teacher" (Connelly & Clandinin, 1988, p. 84). In this case, an explanation of the relationship between the stakeholders would provide a more robust background to the extent of curricular innovation: "Who adopts what, where, when, why, and how?" (Markee, 1992, p. 230). While the iterative nature of curriculum development is acknowledged, the authors seem to overlook the intricate social actions involved in educational change, an essential component for the continuous evolution of curricula.

While these shortcomings exist, they are relatively minor due to the inclusion of references at the end of each chapter, which serve as valuable resources for further reading. The authors demonstrate their extensive knowledge of ELT curricula through their presentation of the topics. Each chapter begins with a vignette and a pre-reading task, followed by an explanation of the topic, a post-reading task, and discussion questions. This structure resembles a lesson plan, enabling teachers to follow along with the chapters easily as if reading their own lessons. It effectively bridges the knowledge or the lack thereof that readers could follow up on. Even if a topic needs to be clarified, it is compensated with relatable examples that are valuable for a practicing teacher who has no time to scour the literature.

Moreover, the book examines curriculum and learners away from their traditional roles in education; instead, it takes a fresh perspective on learners as agents of their own learning and curriculum as a collaborative process amongst the commonplaces in the education system. As such, *curriculum* has many ways to be defined, depending on whom you ask. The book describes it as a cyclical process that involves planning, implementation, and evaluation. The authors situate the language curriculum from a social contextual perspective where teaching and learning happen in the classroom (Graves, 2008; Macalister & Nation, 2011), inspiring teachers to value, reflect, and adapt contemporary teaching methods according to their on-the-ground experiences. Through this, readers are encouraged to continuously update their understanding of the curriculum, assessing the situation in their classrooms, who their learners are, and what they hope to achieve in learning a new language. This inspires readers to have the "nerve to believe that we can make the future what we want" (Eagan, 2003, p. 16) and that there are ways to better language education.

The book critically examines the definition of *learners*, from passive recipients of knowledge to active participants in forming their desired outcomes. The authors mention that these days, the term becomes blurry due to the modes of language acquisition (e.g., online/in-person, AI generative

platforms, and informal/formal schooling) with teachers coming from "different linguacultural backgrounds" (p. x). In other words, the authors do not promise a one-size-fits-all narrative of what English language teachers need to know but put forth possible ways curricula could be designed and implemented across a multitude of cultural and historical contexts. They emphasize the flexibility and constant reflection on the learners' needs, the teachers' beliefs, and institutional practices that embody the curriculum as a text and sociocultural artifact. Whenever a curricular approach fits the learner's needs, teachers can pick the best approach without being limited to pre-arranged options (e.g., top-down curriculum). Teachers know their learners best, and coupled with their observations, needs, and environment analyses, they implement the curriculum accordingly.

Overall, the book comprehensively addresses the essential elements of curriculum development, providing invaluable assistance to novice teachers in navigating the complexities of its context, design, approaches, challenges, and assessment. The book sums up the "basics" of curriculum understanding into a handy "guidebook" that teachers, policymakers, and graduate students can use for a quick reference on handling learners' unique and collective needs. This book is handy and provides teachers with answers to the questions of adopting and implementing any recommended curriculum. The people who will primarily benefit from this book are language teachers who are just about to start a career in teaching. It can be overwhelming for teachers to walk into classrooms full of students from a plethora of cultural backgrounds. It is highly recommended for educators who aspire to make a significant impact in their field, encouraging them to critically reflect on and enhance the delivery and implementation of their language instruction.

References

Cao, Y. (2024, May 17). China's education crossroads: Can the double reduction policy resolve inequalities caused by shadow education? *The British Educational Research Association*.

Christison, M., & Murray, D. E. (2022). *What English language teachers need to know volume III: Design and curriculum* (2nd ed., vol. 3). Routledge. https://doi.org/10.4324/9780429275746-2

Connelly, F., & Clandinin, D. (1988). *Teachers as curriculum planners: Narratives of experience.* Teachers College Press.

Eagan, K. (2003). What is curriculum? *Journal of the Canadian Association for Curriculum Studies, 1*(1), 9–16. https://doi.org/10.25071/1916-4467.16845

Graves, K. (2008). The language curriculum: A social contextual perspective. *Language Teaching*, *41*(2), 147–181. https://doi.org/10.1017/S0261444807004867

Kachru, B. B. (1986). The power and politics of English. *World Englishes*, *5*(2–3), 121–140. https://doi.org/10.1111/j.1467-971X.1986.tb00720.x

Macalister, J., & Nation, I. S. P. (2011). *Case studies in language curriculum design* (1st ed.). Routledge. https://doi.org/10.4324/9780429203763-1

Macalister, J., & Nation, I. S. P. (2020). *Language curriculum design* (2nd ed.). Routledge. https://doi.org/10.4324/9780429203763

Markee, N. (1992). The diffusion of innovation in language teaching. *Annual Review of Applied Linguistics, 13*, 229–243. https://doi.org/10.1017/S0267190500002488

Parent, K. (2011). The teacher as intermediary between national curriculum and classroom. In J. Macalister & I. S. P. Nation (Eds.), *Case studies in language curriculum design: Concepts and approaches in action around the world* (pp. 186–194). Routledge.

Piccardo, E. (2013). Plurilingualism and curriculum design: Toward a synergic vision. *TESOL Quarterly*, *47*(3), 600–614. https://doi.org/10.1002/tesq.110

Learner Corpus Research Meets Second Language Acquisition. Bert Le Bruyn and Magali Paquot. Cambridge Applied Linguistics, 2021. xiii + 275 pp. Approx. ¥6,540. https://doi.org/10.1017/9781108674577

Reviewed by
Trevor Sitler
Kindai University

Researchers of Second Language Acquisition looking for a fresh perspective on learner language, or a fresh perspective on language in general, may want to pick up *Learning Corpus Research Meets Second Language Acquisition*. Still a relatively niche concept in SLA circles, learner corpora are defined in the book's introduction as electronic collections of both oral and written language which was collected in a way that could still be considered authentic and contextualized, like a snapshot of a natural phenomenon. Despite the large amount of learner data available in LCR, the field's influence on SLA has been lacking. Reasons for the gap are varied and are elaborated on by commentary articles in the volume by Sylviane Granger and Florence Myles. However, the two fields do have a large potential to complement each other, with SLA providing a "strong theoretical foundation that is often lacking in LCR studies" and LCR providing "detailed descriptions of interlanguage from a wide range of L1 populations at different proficiency levels" (Granger, 2021, p. 254).

Targeted primarily toward researchers in either one of the fields, the book contains a total of twelve chapters, all of which were reviewed by both an expert in SLA and an expert in LCR. Seven chapters are research articles that attempt to incorporate both SLA theory and LCR methodology. One chapter is devoted specifically to methodological suggestions in corpus analysis. There is also a chapter providing suggestions on how to build a unique, specialized corpus. Finally, there are two commentary chapters from Sylviane Granger, a leading learner corpora researcher, and Florence Myles, an established SLA researcher.

The two commentary chapters by Granger and Myles provide a solid overview of the gap between the two fields as well as their potential to inform each other. Sylviane Granger, one of the leading pioneers of LCR, acknowledges the differences between the two fields while remaining optimistic about the future potential to inform each other. Fundamentally,

these are two separate fields with different research agendas and ways of doing research. SLA, for example, prefers experimental settings, where variables can be controlled and specific language can be elicited. LCR prioritizes naturalistic settings to capture what it deems to be authentic language use. Nonetheless, Granger maintains that LCR can greatly benefit SLA, as illustrated in the book, primarily through studies on L1 transfer and proficiency, which will be discussed in more detail below.

The other commentary article by SLA scholar Florence Myles reconsiders many of the criticisms of LCR she made a decade prior (Myles, 2015). These include a lack of oral corpora, a limited range of tasks, and a lack of thorough documentation in corpus compilation. However, progress has been made since then, especially regarding the increased documentation and more thought-out design of corpora, such as more longitudinal corpora and corpora that are specially designed to fit a specific theoretically informed research question. The increasing amount of oral data is also promising.

Most of the research articles in the book focus on either cross-linguistic transfer or proficiency. Three articles are dedicated to studying L1 transfer through LCR. Ionin and Diez-Bedmar (2021) write about L1 influence in the acquisition of articles from English essays of Russian and Chinese natives. Werner et al. (2021) compare the use of the present perfect among German and Chinese speakers from the Louvain International Database of Spoken English Interlanguage (LINDSEI) (Gilquin et al., 2010). Merilainen (2021) investigates universal tendencies in World English, using data from the above-mentioned ICLE. The study analyzes embedded inversion and the omission of prepositions in Finnish, German, and Swedish students.

There are four articles that research proficiency and development. Paquot et al. (2021) used corpus linguistics methodology to measure phraseological development in a longitudinal corpus. Tracey-Ventura et al. (2021) used a small, specialized corpus to study the individual lexical retention and attrition of 56 participants after study abroad. Verspoor et al. (2021) take an even more microscopic view and analyze a dense set of written data from 22 people using the complex dynamic systems theory (CDST). Polio and Yoon (2021), propose a new measure for accuracy through the analysis of n-grams, or commonly repeated phraseological units, from a corpus of 139 argumentative essays.

The two remaining articles give suggestions for those interested in doing LCR research. Bell et al. (2021) detail the challenges of creating a specialized corpus; Wulff and Gries (2021) provide methodological suggestions with a sophisticated statistical model called MuPDAR (multi-factorial

prediction and deviation analysis using regression/random forest), which can predict the linguistic choices a native speaker would make if they were in the same place as the learners.

This book was intended to convince researchers of either SLA or LCR of the potential contribution from the other field. While the writing style of the articles does contain a fair amount of jargon from both fields and may prove difficult for complete novices of either one, the two commentary articles from Granger and Myles help to put these articles in a broader context. Furthermore, researchers who have written their own articles on the topics of cross-linguistic influence and proficiency may find the new LCR approach to provide a refreshing new perspective to these long-time SLA topics.

All the articles are clear about their theoretical underpinnings, some starting from an explicit SLA framework, some from an LCR framework, and others somewhere in between. Werner et al.'s (2021) research on L1 transfer explicitly uses the framework of Jarvis's (2000) unified model for linguistic transfer. In this framework, there must be both a test for homogeneity for learners of a common L1 background as well as a test for heterogeneity among learners of different L1 backgrounds. This data can be easily provided by large-scale, multi-language corpora. These large data sets also allowed the researchers to perform statistical analysis unique to corpus linguistics, in this case, the above-mentioned MuPDAR. This allowed for a more nuanced and gradient response to the cross-linguistic transfer of present participles.

Similarly, Vespoor et al. (2021) ground their perspective in complex dynamic systems theory (CDST), which claims that language unfolds at different speeds, in different ways, with different individuals (De bot et al., 2007). This framework requires extensive data on individual learners over time, which could be provided by the longitudinal corpus of 22 people that was specifically designed for the study. Through the dense data in the corpus, the study could trace the individual development of written essays and show how the development differed among individuals.

Other articles start with an LCR framework and then bring fresh light on SLA concepts, such as accuracy and proficiency. Polio and Yoon (2021) and Paquot et al. (2021) use the common corpus linguistics statistical technique of collocation to provide a new perspective on accuracy and complexity. Collocations are a measure of the statistical association between words and phrases. For example, "strong coffee" has a stronger statistical association than "robust tea", which sounds a little strange to native speakers. Polio and Yoon used this technique to find unnatural

or awkward combinations of words that may be grammatically correct but still sound off. They then compared these off-sounding collocations to their frequency in a reference, native corpus. This style of measurement allows for a more gradient response as well as the identification of awkward-sounding combinations. Similarly, Paquot et al. (2021) measured the statistical association between verb + direct object structures as a way to track the development of phraseological complexity in a longitudinal corpus. Like Polio and Yoon above, the phraseological collocations were referenced to a native corpus. Learner corpora's large data allows for this level of generalization between word associations, as well as the ability to generalize and represent a native speaker norm that is not tied to any one individual but is instead abstracted from a large population.

The volume also provides a very good summary of the types of corpora that are available for anyone wanting to try their hand at LCR. Corpora such as the LINDSEI (Gilquin et al., 2010), and the ICLE (Granger et al., 2020) are versions of what Granger calls "all-purpose corpora" (2021, p. 246), as they offer the advantage of size and representativeness. However, these specific corpora are not as accessible as one would like. The LINDSEI requires contacting the University of Louvain for access, and the ICLE would cost about 120 euros for a one-year license. Freely accessible all-purpose corpora, such as The International Copus Network of Asian Learners of English (ICNLAE) (Ishikawa, 2023) or the National Institute of Information and Communications Technology (NICT-JLE) (Izumi et al., 2004), are unfortunately not represented in the volume. These corpora may be a smaller barrier of entry for potential LCR initiates.

However, such an "off the peg" corpus, as Myles (2021, p. 265) puts it, may not fit a more specialized research question, in which case, Bell et al. (2021) provide a thorough analysis of the process of how to create one's own corpus. Issues range from how to choose an appropriate task that will engage the student's interest to the mountain of transcription work required for even modest-sized oral corpora. While this may prove daunting, articles such as Vespoor et al. (2021) successfully showcase how corpora tailored to a specific research question and theoretical framework can bolster results.

The book successfully shows the potential for the fields of SLA and LCR to be mutually beneficial. The large datasets of LCR can provide the ability to generalize as well as perform sophisticated statistical techniques like MuPDAR. Likewise, the extensive theories in SLA can help to theoretical ground LCR research, which has been criticized for producing mere descriptions of language with no theoretical foundations.

With the rapid developments in technology in recent years and the increasing availability of oral data and longitudinal corpora, there is a high chance LCR will increase its influence on SLA studies. Publicly available corpora such as the above-mentioned ICNALE, as well as easily accessible corpus tools like Antconc (Anthony, 2024) and Sketch Engine (Kilgarriff et al., 2014) allow anyone with even a minimum amount of computer skill to begin corpus research. While technical at times, this book provides a great introduction to SLA researchers who are looking for a fresh perspective on learner language.

References

Anthony, L. (2024). AntConc (Version 4.3.1) [Computer software]. Waseda University. https://www.laurenceanthony.net/software/AntConc

De Bot, K., Lowie, W. M., & Verspoor, M. H. (2007). A dynamic systems theory approach to second language acquisition. *Bilingualism: Language and Cognition, 10*(1), 7–21. https://doi.org/10.1017/S1366728906002732

Gilquin, G., De Cock, S., & Granger, S. (2010). *The Louvain International Database of Spoken English Interlanguage: Handbook and CD-ROM*. Presses universitaires de Louvain. https://doi.org/10.1111/j.1540-4781.2012.01412.x

Granger, S., Dupont, M., Meunier, F., Naets, H., & Paquot, M. (2020). The International Corpus of Learner English (Version 3). Presses universitaires de Louvain. https://dial.uclouvain.be/pr/boreal/object/boreal:229877

Ishikawa, S. (2023). *The ICNALE guide: An introduction to a learner corpus study on Asian learners' L2 English.* Routledge. https://doi.org/10.4324/9781003252528

Izumi, E., Uchimoto, K., & Isahara, H. (2004). The NICT JLE Corpus: Exploiting the language learners' speech database for research and education. *International Journal of the Computer, the Internet and Management*, *12*(2), 119–125.

Jarvis, S. (2010). Methodological rigor in the study of transfer: Identifying L1 influence in the interlanguage lexicon. *Language Learning, 50*(2), 245–309. https://doi.org/10.1111/0023-8333.00118

Kilgarriff, A., Baisa, V., Busta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., & Suchomel, V. (2014). The Sketch Engine: Ten years on. *Lexicography*, *1*, 7–36. https://doi.org/10.1007/s40607-014-0009-9

Myles, F. (2015). Second language acquisition theory and learner corpus research. In S. Granger, G. Gilquin, & F. Meunier (Eds.), *The Cambridge handbook of learner corpus research* (pp. 309–332). Cambridge University Press. https://doi.org/10.1017/CBO9781139649414.014

Language Support for Immigrants in Japan: Perspectives From Multicultural Community Building. Keiko Hattori, Makiko Shinya and Kurie Otachi (Eds.). Lexington Books, 2023. xvi + 196 pp. ¥7,410. https://doi.org/10.5771/9781666910223

Reviewed by
Karen Masatsugu
Kwassui Women's University

As of the end of June 2024, the total number of foreign residents in Japan was 3,588,956, an increase of 5.2% from the previous year (Immigration Services Agency, 2024), and Japan will reportedly need more than 6 million foreign workers by 2040 according to the Japan International Cooperation Agency (The Japan Times, 2024), creating an urgent need for Japanese language education to enable foreigners to function fully as members of Japanese society. This timely book is a collection of papers about Japanese language education for immigrants in Japan, written by Japanese academics, teachers, and volunteers for an international English-speaking audience. The book is divided into two parts, "History of Immigration Policy and Language Supports" (sic), and "Practice in Various Fields". The two chapters in Part One provide the historical background to Japan's lack of immigration policy and haphazard approach to the provision of Japanese language education for immigrants. The eight chapters in Part Two each give descriptive accounts of grassroots attempts to cater to the needs of the diverse groups of immigrants throughout Japan from an equable and inclusive multicultural community building perspective.

In their introduction, editors Keiko Hattori, Makiko Shinya, and Kurie Otachi explain that this perspective comes from a discourse that differentiates community-based Japanese language classes from conventional Japanese language schools, the former "being seen as places where attendees of equal status gather to discuss, foster mutual understanding, and learn together about local issues" (xi), rather than classrooms focused purely on language learning. These community-based classes have historically relied on resident volunteers, both Japanese and foreign. In Chapter 1, "Japan's policies for Accepting Immigrants and the History of Official Japanese Language Education", Katsuichiro Nunoo gives a comprehensive overview of immigration policies and Japanese language education for immigrants since the 1950s, including a critical analysis of the 2019 Act on

Promotion of Japanese Language Education, the first unified national policy for Japanese Language education.

Izumi Yamada, in Chapter 2, "Japanese Language Learning Support Activities by Local Residents for Immigrants", acknowledges the ambiguity of the term "multicultural symbiosis" (多文化共生, 'tabunka kyousei') and states that Japan should be aiming for "equal multicultural symbiosis: A form of social participation in which the indigenous cultural majority and the new cultural minority participate equally on equal footing" (p. 26). Yamada sees community-based Japanese classes as having two roles and purposes: "mutual learning aimed at social change (as adult education)" and "second language acquisition for social participation (as compensatory education)" (p. 27), in which the following eight chapters present examples of language support in various contexts throughout Japan.

In Chapter 3, "Roles and Practices of Local International Associations", Takashi Yamanoue considers the role of local international associations focusing on Toyonaka in Osaka, in particular how the aim of its Japanese language education program changed following the Great Hanshin Earthquake in 1995 and the realization of the necessity for immigrants to be connected to the community. He details the development of a program that instead of focusing on textbook Japanese, gave the immigrants and their Japanese supporters the opportunity to interact and use Japanese to talk about everyday issues that were relevant to them, "turning the classroom into a community" (p. 49), meeting the self-professed needs of the immigrants.

Chapter 4, "People Involved in Language Learning Support in Community-based Japanese classes", and Chapter 5, "Japanese Language Support for Immigrants in Rural Areas", both use qualitative research data to illustrate the kinds of support being provided. In this chapter, Kurie Otachi and Keiko Hattori focus on three human resource groups involved in community-based classes: language program coordinators, Japanese volunteer supporters, and immigrant supporters, using questionnaires and interviews to investigate each group's perceptions of their roles. In Chapter 5, Keiko Hattori and Makiko Shinya use ethnographic observations to look at the challenges of providing support in rural areas where there is no institutionalized language support. This chapter compares their efforts to establish language support in a prefecture in the Kinki region and in Shikoku, highlighting some important issues that need to be addressed as more immigrant workers are being sent to rural areas (Ministry of Health, Labor and Welfare, 2025).

Chapters 6 and 7 both describe support programs originally initiated by immigrants themselves. In Chapter 6, "Japanese Language Education on Unrecognized "Refugees" in Japan", Shin Matsuo looks at the complicated situation of refugees in Japan and attempts by Villa Education Center, a volunteer group, to provide language support for the Myanmar community in the Takadanobaba area of Tokyo. Matsuo gives a detailed description of the development of the program between 2014 and 2021, showing how participatory learning helped develop both Japanese acquisition and self-affirmation among the participants. In Chapter 7, "Japanese Language Learning for Technical Intern Trainees from Vietnam", Jotaro Kato gives an equally detailed account of providing support to Vietnamese immigrants at Kawaguchi Catholic church, in Kawaguchi city near Tokyo, describing the changes in the Vietnamese community and their needs, from the boat people refugees and their Japan-raised children in the 1990s to the students and technical intern trainees of recent years. Chapter 8, "Challenges and Possibilities of Literacy Education for Immigrants", addresses the difficulty of acquiring literacy in Japanese because of the complicated writing system, and proposes using the Kanji for Everyday Life program devised by the chapter authors Makiko Shinya, Keiko Mikogami, and Aimi Shinjo, for immigrants in Osaka. The program prioritizes learning kanji that immigrants need immediately in daily life, empowering them and enriching their quality of life.

In Chapters 9 and 10, the authors consider the role of night schools in making up for the lack of Japanese language education for immigrants in the Japanese education system. In Chapter 9, "Japanese Language Education for Young Immigrants Who Are Beyond School Age", Tomoko Takahashi investigates two Filipino students who came to Japan after completing junior high school in the Philippines, and were able to enter senior high school after attaining a junior high school diploma at a night school in Osaka. Takahashi discusses the hurdles facing high school-age immigrants and argues that Japanese language education in senior high school needs improvement. Chapter 10, "Literacy Practices Ensuring Education for Resident Koreans in Japan", centers on a 2005 case study of a public night school. Yohei Tanada provides a history of night schools, emphasizing their importance for Korean women who had been marginalized because of both ethnicity and gender. The night school curriculum caters to the students' experiences, backgrounds and needs, enabling them to feel positive about their identity and to participate in Japanese society. The editors conclude the volume by noting that despite recent government legislation, im-

migrants are still regarded as guests in Japan, and by calling for a "mutual transformation that also transforms the majority" (p. 182).

While each of the eight chapters in part two focus on different communities, there are common threads that run throughout the book, notably the tension between the authors' belief in community-based language classes and the aims of the 2019 Act on Promotion of Japanese Language Education, which clarified the responsibilities of national and local governments, and employees in providing Japanese language education to immigrants. Both Nunoo and Yamada welcome the Act, but with caveats. Nunoo notes in Chapter 1 that the Act does not state that foreign nationals have the "right" to learn Japanese, and doesn't define "symbiotic society" (p. 16). He also regrets that the Act doesn't include any mention of the use of plain Japanese, simplified Japanese which is easier for non-Japanese-speakers to understand. Yamada, in Chapter 2, believes the community-based classes are still essential in order to give the local community the chance to interact with immigrants, but sees the Act and the 2020 "Report on the Qualifications of Japanese Language Teachers" as a positive development, with its increased training and accreditation for teachers, "professionalizing Japanese language teachers in several fields, including community-based Japanese language education" (p. 30).

Yamada also points out other problematic underlying issues affecting immigrants, such as not being eligible to vote, and immigrant children not being legally required to attend school, stressing the need for multicultural education in Japanese schools, and for projects which support immigrant children's heritage language and culture. Tokunaga (2018) states that it is rare for public schools to provide education that affirms the native languages, cultural traditions and ethnic identity of immigrant students.

In Chapter 8, the authors acknowledge that while literacy is important in enabling social participation and improving educational and occupational choices, many immigrants, especially in rural areas, have neither the time or opportunity to attend Japanese classes, and that community-based classes usually focus on speaking and listening skills. As Nunoo details in Chapter 1, skilled workers wishing to work in nursing or the care industry need sufficient literacy skills to take written exams in Japanese in order to obtain the required qualifications.

Tanada, in Chapter 10, notes the changes in night school attendance, with the number of resident Korean students falling from over 50% in 1990 to 5% in 2021, while the number of newcomer foreign students has risen to 70%. This echoes Tokunaga's (2018) belief that night schools

are critical alternative educational sites and an important safety net for immigrant students, providing a safe and comfortable setting as Tanada described.

According to the Ministry of Health, Labor and Welfare, the number of foreign workers in Japan reached a record 2,302,587 as of the end of October 2024, an increase of 12.4% since the previous year (Ministry of Health, Labor and Welfare, 2025). Increases in rural areas were notable, with Nagasaki Prefecture recording the highest proportionate increase of 28.1%. Issues faced by Hattori and Shinya in rural areas (Chapter 5) included a lack of coordination between departments in local governments, the administrators not understanding businesses' needs, and not being aware of resources in the area. The authors suggest bottom-up cooperation with local governments and businesses to raise awareness of community-based Japanese learning.

For example, the largest group of foreign workers are Vietnamese, nearly a quarter of the total number of foreign workers nationwide. In his study of Vietnamese immigrants in Chapter 7, Kato is critical of the Technical Intern Training Program, finding in interviews with trainees that they lacked the time or opportunity to attend classes, and that abuse of the trainees was partly due to their lack of Japanese ability.

The research methods presented in the chapters vary in methodology and academic tone but for the most part are accessible to the non-specialist reader, providing detailed descriptive accounts, although some could benefit from tighter editing. Textual errors and possible mistranslations in some sections make these sections hard to understand; for example, in the argument for the continued need for community-based language support since the introduction of the Act for the Promotion of Japanese Language Education in chapter two. Some readers may desire more critical discussion of the use of 'multicultural symbiosis' as a translation of 多文化共生, (*tabunka kyousei*) and how it differs to multicultural coexistence. While Yamada gives four examples of multicultural symbiosis (slave-like, assimilative, equal and colonial) there is no attempt to situate this definition within the literature. Graburn and Ertl (2008) define symbiosis as "living together side-by-side in a relationship, positively" (p. 8), and describe a continuum ranging from being equal but separate to mutual dependence, the latter as evidenced by cooperation between immigrant and local communities after the Hanshin earthquake.

However, these issues do not detract from the authors' message of the need for multicultural community building. The book is important in bring-

ing Japanese insiders' views to an English-speaking audience, and will be of interest to students of Japanese society, immigration and language policy, as well as to those involved with Japanese language education. The writers are Japanese language teachers, experts and volunteers with hands-on experience, and the varied content reflects the diversity and complexity of the immigrant community in Japan. Each chapter raises pertinent issues that urgently need addressing given changes to the rules for specified skilled workers which allow for the possibility of long-term work in Japan, and increased opportunity to bring family members to Japan (Ministry of Foreign Affairs, 2024). The book includes recent data and explains policy changes affecting immigrants up to 2022. Taken as a whole, the book provides a rich description of attempts to provide language support to immigrants in Japan through community-based language classes and night schools, the challenges faced and still to be overcome as Japan increasingly depends on immigrant labor, and provides a convincing argument for multicultural community building.

References

Graburn, N., & Ertl, J. (2008). Internal boundaries and models of multiculturalism in contemporary Japan. In N. Graburn, J. Ertl, & K. Tierney (Eds.), *Multiculturalism in the new Japan* (pp. 1–31). Berghahn Books. https://doi.org/10.1515/9780857450258-002

Ministry of Foreign Affairs. (2024). *What is the SSW?* https://www.mofa.go.jp/mofaj/ca/fna/ssw/us/overview/

Ministry of Health, Labor and Welfare. (2025, January 31).「外国人雇用状況」の届出状況まとめ（令和6年10月末時点）*[Summary of "employment status of foreign nationals" (as of the end of October 2024)].*

Reuters. (2024, July 4). Japan faces shortage of almost a million foreign workers in 2040. *The Japan Times.* https://www.japantimes.co.jp/news/2024/07/04/japan/japan-foreign-worker-shortage/#:~:text=At%20the%20current%20pace%2C%205.91,for%20talents%2C%20experts%20have%20said.

Tokunaga, T. (2018). *Possibilities and constraints of immigrant students in the Japanese educational system*. Paper commissioned for the 2019 Global Education Monitoring Report, Migration, displacement and education: Building bridges, not walls. UNESCO. https://www.jica.go.jp/Resource/jica-ri/publication/other/l75nbg000010mg5u-att/Background_Tokunaga.pdf

JALT Journal Aims and Scope

JALT Journal is a bi-annual, Scopus-approved research journal of the Japan Association for Language Teaching (全国語学教育学会). JALT's larger mission is to support the research programs and professional development of JALT members, promote excellence in language learning, teaching, and research, and provide opportunities for those involved in language education. In line with this mission, *JALT Journal* publishes high-quality English- and Japanese-language, quantitative and qualitative, theoretically-informed and empirically-grounded studies of relevance to second/foreign language education in Japan. Although emphasis is placed on the Japanese context, *JALT Journal* values contributions which also transcend geographical boundaries to illuminate the complex interaction between language, language use, people, education, and society across cultural and socio-political contexts.

When possible, submissions to *JALT Journal* should aim to be both descriptive (*What is my data?*) and explanatory (*Why is my data like this and not otherwise?*) in purpose, and further stimulate scholarly debate, to hopefully improve existing applied linguistic scholarship around the world. Areas of interest include but are not limited to the following:

- Bilingualism and multilingualism
- Classroom-based language education
- Cognitive linguistics
- Contrastive linguistics
- Conversation/discourse/critical discourse analysis
- Critical language pedagogy
- Curriculum design and teaching methods
- Intercultural communicative competence
- Language acquisition/learning
- Language policy and planning
- Language testing/evaluation
- Phonetics
- Pragmatics
- Psycholinguistics
- Semantics
- Sociolinguistics
- Syntax
- Teacher training
- Translation and interpretation
- Vocabulary

Guidelines

Style

Authors are encouraged to submit manuscripts in five categories: (1) full-length articles, (2) short research reports or papers addressing specific theoretical and/or methodological issues in applied linguistics research (*Research Forums*), (3) theory-grounded essays which may include analysis of primary or secondary data (*Perspectives*), (4) comments on previously published *JALT Journal* articles (*Point-to-Point*), and (5) book/media reviews (*Reviews*) either requested by the Book Reviews editor or suggested by the author. Articles should be written for a general audience of language educators; therefore, statistical techniques and specialized terminologies must be clearly explained and their use clearly justified.

Authors are responsible for obtaining permissions for any copyrighted material included in their manuscript submission. When submitting a manuscript based on a thesis or dissertation published in an institutional repository, authors are responsible for checking with the copyright holder for permission to publish portions of the original text.

JALT Journal follows the *Publication Manual of the American Psychological Association, 7th edition* (available from <https://apastyle.apa.org/products/publication-manual-7th-edition>). A downloadable copy of the *JALT Journal* style sheet is also available on our website at <https://jalt-publications.org/jj/>. Instructions for online submissions can also be found at this website.

Format

Full-length articles must not be more than 8,000 words, including references, notes, tables, and figures. *Research Forum* submissions should not be more than 4,000 words. *Perspectives* submissions should not be more than 5,000 words. *Point-to-Point* comments on previously published articles should not be more than 1,000 words in length, and *Reviews* should generally be around 2,000 words. All submissions must be word processed in A4 or 8.5" x 11" format with line spacing set at 1.5 lines. **For refereed submissions, names and identifying references should appear only on the cover sheet**. Authors are responsible for the accuracy of references and reference citations and for obtaining any permissions for copyrighted material contained in the manuscript.

Submission Procedure

Please submit the following two documents in MS Word format to the appropriate editor indicated below:

1. Cover sheet with the title and author name(s), affiliation(s), and contact information of the corresponding author.
2. A blinded manuscript, including title, abstract, and keywords, with no citations referring to the author. Do not use running heads and do not include your biographical information. Follow the *JALT Journal* style sheet. Blinded manuscripts should not include meta-data.

If the manuscript is accepted for publication, a Japanese translation of the abstract will be required. Authors will also be asked to provide biographical information. Insert all tables and figures in the manuscript. Do not send them as separate files.

Submissions will be acknowledged within 1 month of their receipt. All manuscripts are first reviewed by the editorial team to ensure they comply with *JALT Journal* Guidelines. Those considered for publication are subject to blind review by at least two readers, with special attention given to (1) compliance with *JALT Journal* Editorial Policy, (2) the significance and originality of the submission, and (3) the use of appropriate research design and methodology. The first round of review is usually completed within 3 months. Each contributing author of published articles and *Book Reviews* will receive one complimentary copy of the *Journal* and a PDF of the article (*Book Reviews* are compiled together as one PDF). *JALT Journal* does not provide off-prints. Contributing authors have the option of ordering further copies of *JALT Journal* (contact JALT Central Office for price details).

Restrictions

Papers submitted to *JALT Journal* must not have been previously published, nor should they be under consideration for publication elsewhere. *JALT Journal* has First World Publication Rights, as defined by International Copyright Conventions, for all manuscripts published. If accepted, the editors reserve the right to edit all copy for length, style, and clarity without prior notification to authors. Plagiarism, including self-plagiarism, will result in articles not being published or being retracted and may also result in the author(s) being banned from submitting to any JALT publication. For further information, see the JALT Publications Statement of Ethics and Malpractice at: https://jalt-publications.org/jalt-publications-statement-ethics-and-malpractice

Full-Length Articles, Research Forum, Perspectives, and Point-to-Point Submissions

Please upload submissions in these categories to the following website:

https://jalt-publications.org/content/index.php/jj

Manuscripts should follow the **American Psychological Association (APA) 7th Edition style**. Please indicate if your submission is a 1) Full-length article; 2) *Perspectives* article; 3) *Research Forum* article or 4) *Point-to-Point* submission.

For any general inquiries about English-language submissions, please contact:

Jeremie Bouchard, *JALT Journal* Editor
jaltpubs.jj.ed@jalt.org

Japanese-Language Manuscripts

JALT Journal welcomes Japanese-language manuscripts on second/foreign language teaching and learning as well as Japanese-language reviews of publications. Submissions must conform to the Editorial Policy and Guidelines given above. Authors must provide a detailed abstract in English, 500 to 750 words in length, for full-length manuscripts and a 100-word abstract for reviews. Refer to the Japanese-Language Guidelines (following page) for details. Please send Japanese-language manuscripts to:

Masayuki Kudo, *JALT Journal* Japanese-Language Editor
jaltpubs.jj.ed.j@jalt.org

Reviews

The editors invite reviews of books and other relevant publications in the field of language education. A list of publications that have been sent to JALT for review is published bimonthly in *The Language Teacher* and can be found online in each issue at <https://jalt-publications.org/tlt/>. Review authors receive one copy of the *Journal*. *JALT Journal*'s latest *Reviews* guidelines can be found in Melodie Cook's *Expositions* piece published in the November 2023 issue. Please send submissions, queries, or requests for books, materials, and review guidelines to:

Melodie Cook, *JALT Journal* Reviews Editor
jaltpubs.jj.reviews@jalt.org

Inquiries about Subscriptions, Ordering *JALT Journal*, or Advertising

JALT Central Office
Marunouchi Trust Tower Main Building 20F
1-8-3 Marunouchi, Chiyoda-ku, Tokyo 100-0005 JAPAN
Tel.: (+81) 3-5288-5443
Email: jco@jalt.org URL: https://jalt.org

日本語論文投稿要領

*JALT Journal*では日本語で執筆された (a) 論文、(b) 研究報告、(c) 展望論文、(d) *JALT Journal*に掲載された著作物へのコメント・考察、(e) 書評を募集しています。(a) 論文と (b) 研究報告の違いは、以下の通り字数制限による違いです。(c) 展望論文は、言語教育研究に関する課題に焦点をあてた短い論文で、先行研究の検証、理論や1次2次データに基づく議論などを含むものです。文体:一般的な学術論文のスタイルを用い、章立ての仕方や参考文献のデータの書き方などは、*Publication Manual of the American Psychological Association* (7th edition) の定める方式に合わせて下さい。*JALT Journal*書式シート (日本語原稿用) を以下からダウンロードできます<https://jalt-publications.org/jj/>。なお、*JALT Journal*の読者は現場の教師が主なので、特殊な専門用語や統計的手法は、わかりやすく定義するか説明を加えるなどして下さい。原稿:長さは、参考文献リストも含め、(a) 論文は25,000字、(b) 研究報告は13,000字、(c) 望論文は16,000字、(d) *JALT Journal*に掲載された著作物へのコメント・考察は2,000字、(e) 書評は1,500~3,000字以内です。A4の用紙に横書きで、1行40字、1ページ30行で印刷して下さい。手書きの原稿は受け付けません。

提出するもの:
*JALT Journal*書式シート(日本語原稿用)を参考に作成の上、電子メールの添付書類でお送りください。なお、上記(a)論文~(e)書評のどのカテゴリーへの投稿かを明記ください。審査を経て掲載の認められた草稿は、図表などを全て写植版にしたものにして提出願います。

査読: 編集委員会で投稿要領に合っているかどうかを確認したあと、少なくとも二人の査読者が査読を行います。査読者には執筆者の名前は知らされません。査読の過程では特に、原稿が*JALT Journal*の目的に合っているか、言語教育にとって意味があるか、独創性はあるか、研究計画や方法論は適切か等が判定されます。査読は通常二か月以内に終了しますが、特に投稿の多い場合などは審査にそれ以上の時間がかかることがあります。

注意: *JALT Journal*に投稿する原稿は、すでに出版されているものや他の学術雑誌に投稿中のものは避けて下さい。*JALT Journal*は、そこに掲載されるすべての論文に関して国際著作権協定による世界初出版権を持ちます。なお、お送りいただいた原稿は返却しませんので、控を保存して下さい。

投稿原稿送り先またはお問い合わせ:

〒001-0016 北海道札幌市北区北16条西2丁目 藤女子大学
JALT Journal 日本語編集者　工藤 雅之
電話: 011-736-5368
jaltpubs.jj.ed.j@jalt.org

JALT Journal 第 47 巻　第 1号

2025年 4月20日　印刷
2025年 5月1日　発行
編集人　ブシャー　ジェレミ
発行人　クレア・カーネーコー
発行所　全国語学教育学会事務局
〒100-0005 東京都千代田区丸の内1-8-3 丸の内トラストタワー本館20階
TEL (03) 5288 5443
印刷所　コーシンシャ株式会社
〒530-0043　大阪市北区天満1-18-4天満ファーストビル301 TEL (06) 6351-8795

JALT2025

全国語学教育学会
第51回年次国際大会教材展示会
2025年10月31日～2025年11月2日
国立オリンピック記念青少年総合センター

51st Annual International Conference on Language Teaching and Learning & Educational Materials Exhibition

October 31 – November 2, 2025

National Youth Olympic Center, Yoyogi, Tokyo, Japan

https://jalt.org/conference/jalt2025