ISSN 2760-0904

Pan-Pacific Association of Applied Linguistics

Selected Papers of the 29th Conference of Pan-Pacific Association of Applied Linguistics

PAAL 2025 International Conference

New Frontiers in Language Education and Applied Linguistics

August 20–21, 2025

Juntendo University, Tokyo, Japan

Contents

Empathy as Communication Pedagogy: Adapting ECCS and SPIKES to Support Neurodiverse Learners

Yue Shi Yang

National Chung Cheng University

lngben113@alum.ccu.edu.tw

Abstract

This case study adopts an applied linguistics perspective to investigate the practical application of empathetic communication in educational and social settings involving a young adult with ADHD. It addresses the lack of structured communication models for educators interacting with neurodiverse learners. Through a detailed qualitative analysis of the participant's narrative (Y), the research explores her lived experiences with academic marginalization and subsequent self-stigma. The study centers on the adaptation of two established frameworks, the Empathetic Communication Coding System (ECCS) and the SPIKES protocol, transforming them into a pedagogical model for teacher-student and peer-to-peer interactions. Findings illustrate how formalized and informal empathetic acts such as teachers granting learning accommodations and actively validating struggles, are crucial communication strategies that diminished Y's self-stigma and fostered a positive self-image. The adapted model provides educators with specific, reproducible techniques, including reflective listening and structured invitations to speak, to navigate sensitive topics with neurodiverse students. This paper contributes a valuable, communication-based pedagogical framework for fostering supportive learning environments and advancing applied linguistics research in educational equity.

Keywords

Communication strategies, Empathy, ADHD, Educational settings, Pedagogical framework

1 Introduction

Empathy, defined by Rogers (1957) as the profound ability to grasp another's feelings and experiences "as though you were experiencing their situation," is not merely a psychological construct but a powerful sociolinguistic tool essential for effective communication (Batson, 1991). While its role in caregiving and healthcare communication is well-documented (Chu et al., 2024; Lu et al., 2017), its function as a pedagogical strategy in daily social and educational settings remains critically underexplored within applied linguistics. For individuals with conditions like ADHD and associated learning difficulties, stigma and academic failure

frequently diminish self-esteem. This challenge highlights an urgent need for research into how structured communication, rather than clinical intervention alone, can facilitate their linguistic and social integration in learning environments.

This study responds to this gap by focusing on the adaptation and application of two established theoretical frameworks: the Empathetic Communication Coding System (ECCS) (Bylund & Makoul, 2005) and the SPIKES protocol (Baile et al., 2000). The ECCS, originally designed for analyzing emotional validation in medical interviews, is repurposed here to examine teacher-student dialogue, while the SPIKES framework, a structured approach for delivering difficult news, is modified to guide educators in navigating sensitive accommodation discussions. By exploring these approaches, we aim to transform empathy into a replicable communication model applicable to educational contexts, fostering a supportive and inclusive atmosphere. In a high-pressure educational system, such as the one prevalent in Taiwan where academic achievement is heavily tied to social identity and conformity, this communicative scaffolding is particularly crucial for counteracting unique societal pressures that intensify feelings of self-stigma among neurodiverse learners.

Employing a qualitative case study approach, this research centers on the in-depth interview narrative of one young adult (Y), whose journey navigating higher education in a language-centered discipline provides rich data on communication breakdown and repair. The study's purpose is to move the discourse beyond a clinical, deficit-oriented model of ADHD to emphasize the condition's discursive and socio-emotional dimensions. Specifically, we examine the communication styles of individuals with ADHD and how the deliberate application of empathetic communication strategies by educators and support networks is associated with the participant’s ability to diminish self-stigma, improve resilience, and foster the development of a positive self-identity.

This research significantly contributes to the field of applied linguistics and communication pedagogy by addressing two guiding questions: 1. How can the adaptation of frameworks like ECCS and SPIKES provide educators with a structured, communication-based pedagogical model to effectively support the socio-emotional needs of students with ADHD? 2. How is empathetic communication associated with a young adult with ADHD building a positive self-identity, and what implications does this have for understanding and enhancing the communication styles of neurodiverse learners?

2 Past Studies

The prior discussion established the urgent need for structured, communication-based pedagogical models to support neurodiverse learners in language-centered environments. This section lays the theoretical groundwork by examining three interconnected areas. First, we situate the challenges of ADHD-related stigma within the framework of discursive analysis in educational settings. Second, we delve into the core principles of the

Empathetic Communication Coding System (ECCS) and the SPIKES protocol, arguing for their significant pedagogical and theoretical value within applied linguistics. Finally, we explore the existing literature supporting empathy not just as an emotion, but as a demonstrable communication strategy essential for student identity and resilience.

2.1 ADHD, Stigma, and Discursive Barriers in Educational Settings

The challenges faced by individuals with ADHD extend beyond clinical symptoms, manifesting significantly in their socio-emotional and communication experiences within educational settings. From an applied linguistics perspective, this psychological distress is often emphasized by discursive barriers encountered in daily academic interactions. Research highlights that stigma is not merely an internal label but is actively constructed, maintained, or dismantled through daily language use and interactional patterns (Goffman, 1963). In the classroom, subtle linguistic cues from teachers and peers can reinforce a student's identity as "deficient" rather than "neurodiverse," leading to the internalization of negative self-labels (Dahl & Dammeyer, 2024).

2.2 The Pedagogical Potential of Applied Communication Frameworks: ECCS and SPIKES

Furthermore, students with ADHD frequently encounter difficulties in core communication skills relevant to language-centered disciplines, such as turn-taking, maintaining discourse cohesion, and organizing spoken or written narratives. These challenges often lead to misinterpretation, with educators incorrectly attributing these communicative differences to a lack of effort or academic defiance, thereby intensifying marginalization (Chu et al., 2024). While the literature acknowledges the need for academic accommodations, a critical gap remains in providing educators with structured communication protocols designed to proactively mitigate the linguistic and socio-emotional damage caused by stigma in real-time classroom discourse. This framework is essential to move beyond merely diagnosing deficits and toward empowering supportive interaction.

2.3 The Pedagogical Potential of Applied Communication Frameworks: ECCS and SPIKES

To develop a structured pedagogical response, this study adapts two frameworks rooted in communication research: the Empathetic Communication Coding System (ECCS) and the SPIKES protocol. We argue that these models, typically confined to clinical settings, hold significant potential as trainable communication strategies for educational professionals.

The ECCS, developed by Bylund and Makoul (2005), is a fine-grained observational tool originally used to code and quantify physician communication skills, specifically targeting emotional validation, non-verbal cues, and active listening in medical encounters. In the current context, we argue that ECCS's principles, particularly its focus on recognizing and responding to emotion (reflective listening and emotional validation) , are directly transferable to teacher-student dialogue. For an educator, recognizing a student's frustrated

language as a signal of emotional distress (rather than just academic defiance) requires a validated coding and response system, which ECCS can offer.

Similarly, the SPIKES protocol (Baile et al., 2000), a six-step conversation model (Setting, Perception, Invitation, Knowledge, Empathy, Summary), is an invaluable structured dialogue tool for the classroom. Although initially designed for delivering difficult medical news, SPIKES offers a precise blueprint for educators to navigate sensitive academic topics, such as discussing a failing grade, recommending special education services, or adjusting a student's Individualized Education Program (IEP). The "Invitation" step is crucial as it explicitly grants the student agency and control over the conversation's flow, a principle highly valued in studies of learner autonomy in language education (Benson, 2006). By adapting and integrating ECCS and SPIKES, this study provides a new, combined model positioned not as clinical evaluation, but as trainable pedagogical communication strategies for educators.

2.4 Empathy as a Strategy for Identity and Resilience in Education

Psychological research consistently confirms that empathy is fundamental to fostering self-acceptance and emotional repair (Rogers, 1957). However, within pedagogy, empathy functions as a social-affective bridge that connects the instructor's understanding with the learner's sense of belonging, a topic often addressed under teacher affect and rapport in applied linguistics. Lu et al. (2017) emphasize that effective support requires deeply felt empathy, particularly for vulnerable populations. In education, this translates into the teacher's willingness to engage in genuine empathetic labor, such as the conscious management and expression of emotion to foster a positive, non-judgmental classroom environment. When educators deploy deliberate empathetic behaviors, they engage in a powerful act of discursive affirmation. These actions actively counteract the internalization of negative self-labels, enabling students to "reconstruct self-identity" (Batson, 1991) and transform their self-narratives. This process is essential for building resilience, which is often demonstrated through improved willingness to communicate and sustained academic engagement.

2.5 Methodology

This qualitative case study employed a semi-structured interview with the participant, Y, lasting approximately 90 minutes. The session was audio-recorded and transcribed verbatim for analysis. Data were analyzed using a thematic approach, where coding was deductively guided by the ECCS and SPIKES frameworks to identify specific empathetic interactions and their impact on Y's self-identity.

3 Results and Discussions

This research explored how empathy and structured communication influenced the self-identity and social integration of Y, a young adult with ADHD and learning difficulties. Y's narrative consistently highlights the

significant impact of empathetic interactions from peers and educators, not merely as emotional support, but as concrete communicative acts that helped her manage self-stigma and reconstruct a positive self-identity. The findings are presented in two main analytical sections, focusing on how different communication strategies mitigated discursive barriers, followed by a discussion that integrates these results with the broader literature on communication pedagogy.

3.1 Communicative Affirmation: The Power of Informal Pedagogical Empathy

This study significantly broadens the understanding of empathy's influence beyond caregiving contexts (e.g., Chu et al., 2024) by detailing its powerful role in creating inclusive educational discourse communities. Y's narrative demonstrates that informal yet deliberate empathetic actions from educators function as powerful non-verbal communicative acts of academic affirmation.

Y specifically recalls a high school English teacher and a university Spanish professor who provided accommodations like extended test time, helping her feel recognized and valued. This is captured in her statement: "My English teacher in high school allowed me additional time for tests, and that made me feel recognized." (Gazhong ingwen laoshi hui yanchang wode tseyanshijian, zhe shiwo gandao ziji shi beilijie de. 高中英文老師會延長我的測驗時間，這使我感到自己是被理解的。)

This simple action of granting extended time, often viewed as an administrative task, is reinterpreted here as a pragmatic act of validation. Crucially, the participant's narrative indicates that these empathetic behaviors were largely spontaneous acts of individual pedagogical care by the teachers, rather than mandated institutional norms, suggesting that personal communicative agency plays a vital role in student support. It communicates the message: "Your cognitive needs are seen, and your time is valuable, regardless of the typical metric." This form of non-verbal, empathetic communication helped Y move beyond the internal narrative of "being behind" or "deficient." By proactively addressing a systemic barrier, the teacher engaged in a powerful form of anti-stigma discourse, which contributes directly to a sense of belonging and enabling Y to see herself beyond the labels of her diagnoses. This suggests that pedagogical policy can be a form of powerful, positive communication.

Furthermore, Y's interactions with classmates were crucial for establishing a safe discourse environment. She noted: "I felt comfortable sharing my story when friends opened up about their challenges first." (Dang pengyoumen shouxian fenxiang tamende nanchu shi, wo tsai kanjiang zijide gushi. 當朋友們首先分享他們的難處時，我才敢講自己的故事。) This example illustrates the process of discourse community formation where vulnerability and shared narratives served as a foundational communication strategy. When peers initiated self-disclosure (a type of empathetic exchange), they lowered the communicative risk for Y, allowing her to transition from guarded, planned speech to more spontaneous and authentic self-expression. These

exchanges fostered self-acceptance and allowed her to focus on her strengths, such as her achievements in music and language studies, rather than her deficits.

3.1 Structured Dialogue as Pedagogical Scaffolding: Applying ECCS and SPIKES

The research identified clear variations in Y's communication methods, revealing her dependence on structured communication strategies to navigate social interactions. This planning is a compensatory discourse strategy aimed at managing her ADHD-related challenges with spontaneous speech and topic maintenance. She noted that she "plan[s] my words ahead of time to prevent errors" (*Wohui tiqiank kaolu yaoshuode neirong, zheyang bijiao burongyi chuchuo.* 我會提前考慮要說的內容，這樣比較不容易出錯), which stands in contrast to the more fluid, spontaneous communication expected of typically developing individuals. This strategic approach highlights the discursive challenge faced by individuals with ADHD in maintaining interactional coherence.

The study demonstrates how structured empathetic communication, guided by the principles of ECCS and SPIKES, can serve as effective pedagogical scaffolding to bridge this gap. A university professor's use of reflective listening, which is a core ECCS principle, during a discussion about academic performance proved crucial. Reflective listening validates the speaker's emotional state before moving to content. The professor's statement, "It seems like the situation was truly difficult for you. Could you share more details about how you managed it?" (Tinqilai zhezhong qingkuang duini laishuo hen quyou tiaozhanzing, nineng duo fenxiang yixia nishi ruhe yingdui de ma? 聽起來這種情況對你來說很具挑戰性，你能多分享一下你是如何應對的嗎？)

This example is a clinical demonstration of ECCS in practice. The phrase "It seems like the situation was truly difficult for you" explicitly validates Y's frustration (responding to emotion) before asking a follow-up question. This discursive pattern created a safe space that encouraged open dialogue and positioned Y not as a problem student, but as an active agent managing a difficult situation. This pedagogical move affirms her communicative worth and encourages her to engage in detailed self-analysis.

Similarly, the SPIKES framework provided a structured process for handling sensitive conversations, particularly through the crucial "Invitation to Speak" step. This strategy is essential for neurodiverse learners as it reduces anxiety by giving the participant a sense of procedural empowerment and communicative control. The professor's opening, "Do you want to talk about your worries regarding time management in this course?" (Ni xiangyao tantan nizai zhemenke guanyu shijianfenpei shang de danyou ma? 你想要談談你在這門課關於時間分配上的擔憂嗎?)

By asking this question, the professor adheres to the SPIKES structure, which prepares the speaker (Y) for the sensitive topic and explicitly invites her to participate. This contrasts sharply with a direct statement like "We need to talk about your time management." The structured "Invitation" minimizes surprise and anxiety,

allowing Y to deploy her planned communication strategies effectively, thereby reducing the likelihood of communication errors or retreat.

3.2 Discussions

The findings of this case study offer several significant contributions by positioning empathetic communication as a trainable pedagogical protocol rather than a spontaneous emotional trait.

While prior research often focuses on clinical interventions, Y's narrative highlights the profound impact of informal, empathetic actions. The challenge for educators is that informal acts are often inconsistent. This study addresses that by arguing for the formal adoption of ECCS and SPIKES principles to standardize these informal acts into a reliable support system. The ECCS provides the micro-linguistic tools (reflective statements) while SPIKES provides the macro-discursive structure (conversation steps) for complex teacher-student dialogue. By explicitly aligning with the principles of Universal Design for Learning (UDL) which advocates for multiple means of engagement and representation, this combined approach is a crucial contribution to Universal Design for Communication (UDC). It offers educators a tangible framework to address both the emotional and cognitive needs of neurodiverse learners within a standard inclusive pedagogy.

This research directly addresses the call within applied linguistics for methods that enhance educational equity for marginalized groups. The study demonstrates that communication training (using ECCS/SPIKES) can be a powerful intervention tool. By focusing on person-centered methodologies, where validation and structured communication empower individuals to advocate for themselves, this model moves beyond the deficit-oriented model of ADHD. This shift emphasizes the emotional and social dimensions of the condition and underscoring the importance of training educators in specific, evidence-based discursive practices that foster self-affirmation and resilience.

Finally, by presenting Y's experiences within a Taiwanese educational setting, the research illustrates how societal pressures for academic conformity can interact with an ADHD diagnosis to intensify self-stigma. The individualized, narrative-driven perspective of this study suggests that future research should involve classroom discourse analysis to empirically code teacher and peer communicative behaviors using the adapted ECCS framework. This would validate the framework's effectiveness across various cultural and institutional contexts, further solidifying its role as a key communication pedagogy for educational inclusion. The following table (Table 1) shows the difference of the original ECCS and SPIKES clinical focus and the adapted pedagogical application.

Table 1

The differences of original ECCS/SPIKES and the adapted pedagogical application

Framework	Original Clinical Focus	Adapted Pedagogical Application
ECCS	Coding physician empathy in medical interviews.	Validating student distress and using reflective listening in teacher-student dialogue.
SPIKES	Delivering difficult medical news/diagnoses.	Navigating sensitive discussions on academic accommodations and IEP adjustments.

4 Conclusions and Limitations

This case study successfully demonstrated that empathetic communication functions as a critical, trainable pedagogical tool for mitigating the challenges associated with ADHD-related stigma. By adapting and integrating the ECCS and the SPIKES protocol, this research moves beyond conceptualizing empathy as a mere emotional trait and establishes a structured, communication-based model for educators.

The findings illustrate that both informal acts of affirmation (like granting accommodations) and structured discursive techniques (like reflective listening and the "Invitation to Speak") are crucial strategies for teachers to engage in anti-stigma discourse. By empowering the participant, Y, to navigate sensitive academic topics with reduced anxiety and increased control, the model effectively fostered a positive self-identity and reinforced resilience. This study contributes a valuable pedagogical framework to applied linguistics, and emphasize the need of formal communication training to enhance educational equity and support the socio-emotional well-being of neurodiverse learners.

Despite these contributions, this study is subject to several limitations. As a qualitative single-case study, its findings are deeply contextualized by the specific experience of Y and the Taiwanese educational environment; thus, generalizability to other populations or cultural contexts is limited. The reliance on retrospective interview data may also introduce recall bias.

For future research, we propose two key directions aligned with applied linguistics. Conduct classroom discourse analysis in diverse educational settings to empirically validate the adapted ECCS/SPIKES model, coding teacher and student communicative turns to establish the framework’s reliability. Apply this pedagogical model in a larger-scale comparative study involving a control group, measuring the explicit impact of communication training for teachers on student self-efficacy and academic communication willingness.

Acknowledgments

This paper has been previously presented at the 2025 Pan-Pacific Association of Applied Linguistics Conference (PAAL) held at Juntendo University, Tokyo, Japan. The author is deeply grateful to the anonymous reviewers of the conference committee for their rigorous comments and suggestions, which significantly

improved this work. The author also extends gratitude to the Office of International Affairs at National Chung Cheng University, Taiwan, for providing the necessary travel grants and financial support to conduct this research. Additional thanks go to Prof. Victoria Rau, Ms. Neille Yasmin Avinante, and all conference attendees for their insightful feedback. All errors remain the author's own.

References

Baile, W. F., Buckman, R., Lenzi, R., Glober, G., Beale, E. A., & Kudelka, A. P. (2000). SPIKES—A six step protocol for delivering bad news: Application to the patient with cancer. *The Oncologist*, *5*(4), 302–311. https://doi.org/10.1634/theoncologist.5-4-302.

Batson, C. D. (1991). *The altruism question: Toward a social-psychological answer.* Lawrence Erlbaum Associates, Inc.

Benson, G. S. (2006). Employee development, commitment and intention to turnover: A test of 'employability' policies in action. *Human Resource Management Journal*, 16, 173-192. https://doi.org/10.1111/j.1748-8583.2006.00011.x

Bylund, C. L., & Makoul, G. (2005). Examining empathy in medical encounters: An observational study using the Empathic Communication Coding System. *Health Communication*, *18*(2), 123–140. https://doi.org/10.1207/s15327027hc1802_2 .

Chu, N., Pho, J., Dark, L., Tan, A., Alford, S., Tang, C. Y., Ellison, C., & Lim, D. (2024). A scoping review into the service needs of people from culturally and linguistically diverse backgrounds living with disability to engage in meaningful occupations. *Australian occupational therapy journal*, *71*(3), 408–422. https://doi.org/10.1111/1440-1630.12938

Dahl, O., & Dammeyer, J. (2024). Understanding care from the view of individuals with cerebral palsy: pity, politics, and pride. *Health & Social care in the Community. 2024(1), 1-12.* https://doi.org/10.1155/2024/5574535

Goffman, E. (1963). *Stigma. Notes on the management of spoiled identity.* London: Penguin Books.

Lu, H. Y., Liu, C. H. & Tsao, W. L. (2017). "Getting warm with each other" in real and virtual communities: The meaning of social support to caregivers with a family member suffering from dementia. *Communication and Society, 42(3),* 145-188. http://dx.doi.org/10.30180/CS.20170_(42).0003

Rogers, C. R. (1957). The necessary and sufficient conditions of therapeutic personality change. *Journal of Consulting Psychology, 21*(2), 95–103. https://doi.org/10.1037/h0045357

L2 Intelligibility and Extralinguistic Factors: A Case of Japanese-Accented English

Gento Okawa

Simon Fraser University

gento_okawa@sfu.ca

Abstract

This study investigated the effects of extralinguistic factors on the intelligibility of Japanese-accented English (JaE), taking into account speaker exposure to English, listener knowledge of Japanese, and listener first language (L1) status. The study first analyzed the characteristics of JaE with a focus on phonemic substitutions, using the Speech Accent Archive (Weinberger, 2015). Speech stimuli were selected based on the analysis and recorded by three Japanese speakers of English with varied exposure to English. Listeners (n = 18), including second language (L2) speakers of Japanese and L1 and L2 speakers of English, listened to the recordings and completed minimal pair identification tasks. The results revealed a significant effect of speaker exposure on vowel and consonant intelligibility and a positive influence of listener knowledge of Japanese on vowel intelligibility, but not on consonant intelligibility. Listener L1 status showed no significant effect on vowel or consonant intelligibility. Overall, speaker exposure appears to play a primary role in the intelligibility of JaE. The finding of listener knowledge of Japanese might suggest an "extended" interlanguage speech intelligibility benefit (ISIB) on vowels, in which listeners have an advantage in L2 intelligibility when they speak speakers' L1 as an additional language.

Keywords

Intelligibility, Japanese-accented English, L2 perception, L2 comprehension

1 Introduction

Since the seminal work of Munro and Derwing (1995), intelligibility (i.e., actual understanding) has been increasingly popular in research on second language (L2) pronunciation, along with comprehensibility (i.e., ease of understanding) and accentedness (i.e., the difference between an expected and perceived pronunciation). They demonstrated the partial independence of these three constructs, which was also observed in several subsequent studies (e.g., Derwing & Munro, 1997), providing clear evidence that accented speech can be

intelligible and comprehensible. Pedagogical contexts shifted accordingly, prioritizing understanding over nativelike pronunciation (the intelligibility principle; Levis, 2005).

This study aims to contribute to the line of L2 pronunciation research by examining intelligibility, with a focus on English spoken by first language (L1) Japanese speakers and extralinguistic factors. Specifically, this study investigates three extralinguistic factors: speaker exposure to English, listener knowledge of Japanese, and listeners' L1 status.

2 Literature Reviews

2.1 L2 Intelligibility

A growing body of research has examined L2 intelligibility from various perspectives. One line of research has focused on what causes intelligibility breakdown. Jenkins (2002) demonstrated that pronunciation errors, including segmental and suprasegmental errors, are central to intelligibility issues. Especially, Nagle and Huensch (2020) found that phonemic errors (i.e., deletion, insertion, or substitution of a phoneme) were negatively correlated with intelligibility.

Another perspective of interest is listener-based factors on intelligibility. Prior research has examined how listeners' experience affects intelligibility judgments. One of the earliest studies is Gass and Varonis (1984), which provided evidence that listeners' familiarity with general L2 speech and with a particular accent positively influences the judgment of L2 intelligibility. More recently, Pesantez and Lima (2023) revealed that the L2 listeners sharing L1 background with the speakers outperformed the other groups (L1 listeners and L2 listeners with different L1s), revealing benefits of sharing L1s in L2 intelligibility judgments (the matched interlanguage speech intelligibility benefit [ISIB]; Bent & Bradlow, 2003). However, research on the ISIB has mixed results. In fact, Munro et al. (2006) revealed no significant contribution of listeners' L1 linguistic background on intelligibility, and Jułkowska and Cebrian (2015) showed that L1 listeners consistently outperformed L2 listeners.

Although many studies examined listener-based factors, only a few have considered speaker-based factors beyond an utterance level. One exception is Huensch and Nagle (2021). In their research, the speakers were recruited from different levels of language courses, based on which their proficiency was defined. The results showed that proficiency had a positive impact on segmental and suprasegmental errors as well as fluency. This indicates that speakers at different proficiency levels exhibited distinct patterns of linguistic errors that significantly influenced intelligibility judgments.

2.2 The Current Study

Building upon previous research highlighting an essential role of listeners' and speakers' experiences, the present study aimed to examine the intelligibility of Japanese-accented English (JaE) with a focus on three

variables: speaker exposure to English, listener knowledge of Japanese, and listener L1 status (i.e., whether the listener is the L1 speaker of English). Speaker-based extralinguistic variables are relatively understudied despite their important role in intelligibility (e.g., Huensch & Nagle, 2021), which motivated the inclusion of speaker exposure to English as a speaker-based variable. Additionally, the choice of listener knowledge of Japanese was motivated by the effect of listeners' familiarity. Prior research has shown its positive impact on L2 intelligibility judgment, including the ISIB (e.g., Gass & Varonis, 1984; Pesantez & Lima, 2023). However, few have examined listeners' acquired knowledge of speakers' L1, as opposed to a shared L1 background. Finally, listener L1 status was also considered, given that employing both L1 and L2 listeners' judgment better reflects real-life communication (Jenkins, 2002). Accordingly, the following research questions were posited.

- Do speakers' exposure to English affect the judgments of L2 intelligibility?
- Does listeners' knowledge of speakers' L1 affect L2 intelligibility judgments?
- Does listeners' L1 status affect L2 intelligibility judgments?

3 Methods

3.1 Selection and Recording of Speech Stimuli

This section discusses the process of selecting and recording speech stimuli. This study operationalized intelligibility through identification tasks utilizing minimal pairs. This choice was made for practical reasons in administering the experiment. Accordingly, this study first analyzed the characteristics of JaE to decide which minimal pairs to employ in the identification tasks. The following subsections discuss the analysis of JaE (Section 3.1.1), the selection of stimuli (Section 3.1.2), and the recording of speech samples (Section 3.1.3).

3.1.1 Analyzing Japanese-Accented English (JaE)

JaE is often characterized by phonemic substitutions and vowel epenthesis (Watanabe, 2017). To create minimal pair identification tasks, I examined phonemic substitutions using the Speech Accent Archive (Weinberger, 2015). I manually analyzed the phonetic transcriptions of 16 recordings of Japanese speakers of English available on the Speech Accent Archive, and a total of 25 phonemic substitution pairs common across JaE were identified (see Table 1). The left sound in a pair represents a sound that JaE speakers tend to have difficulty producing, while the right sound represents what they tend to articulate instead (e.g., for /i/-/ɪ/, they tend to pronounce these /ðiz/ as [ðɪz]). Many of these pairs have been documented in prior research on JaE, such as tense vs. lax, monophthong vs. diphthong, interdental vs. alveolar fricative, and /l/ vs. /ɹ/ (Aoyama et al., 2023; Nishi, 2002; Watanabe, 2017). Among consonants, voiceless-voiced substitution pairs were often observed in this study, which is less documented in JaE literature.

Table 1

Phonemic Substitutions Characterizing JaE

Vowels	/i/ - /ɪ/	/ɑ/ - /ɔ/	/ɔ/ - /ɑ/	/ɔ/ - /ʌ/	/ɔ/ - /æ/	/eɪ/ - /e/	/oʊ/ - /ɔ/
Consonants	/b/ - /p/	/t/ - /d/	/tʃ/ - /dʒ/	/d/ - /t/	/g/ - /k/	/v/ - /f/	/θ/ - /s/
	/θ/ - /t/	/ð/ - /d/	/ð/ - /s/	/ð/ - /z/	/z/ - /θ/	/z/ - /s/	/ʃ/ - /s/
	/dz/ - /ts/	/ɹ/ - /l/	/l/ - /ɹ/	/w/ - /v/			

3.1.2 Selecting Speech Stimuli

As speech stimuli, minimal pairs were created based on these 25 phonemic substitutions. A minimal pair consists of a target word, which includes a left sound in a pair, and a foil word, which includes a right sound. All the target words in minimal pairs, other than *bathe* and *clothe*[1] were chosen from the COCA's top 5000 entries of Word Frequency (Davis, 2008) to guarantee that the subjects were all familiar with the speech stimuli. Three minimal pairs were created for each phonemic substitution (see Appendix A), amounting to a total of 75 minimal pairs.

To record the speech stimuli in a natural context, three passages were created[2]. Each passage contained 25 target words for each phonemic substitution from the minimal pairs. An effort was made so that the target words would not be subject to linking or assimilation.

3.1.3 Recording Speech Stimuli

To record these passages, three Japanese speakers of English were invited. They are all L1 speakers of Japanese and started learning English in middle school. The first speaker (SP1) had 1.5 years of experience in a language school in the US. The second speaker (SP2) was a fourth-year student in an English-medium instruction program at a Japanese university and had one year of experience in the US as an exchange student. The third speaker (SP3) was a university student in the US and lived there for six years. Accordingly, SP1 was defined as the speaker with the least exposure to English, whereas SP3 as the one with the most.

They were assigned one of the three passages and asked to record it. The recording was conducted in a quiet room with written instructions provided by the experimenter. The audio files were adjusted for peak amplitude and then converted into MP3 files. To create identification tasks, only the target words from the

[1] Given the Word Frequency list includes *bath* and *clothes*, their derived forms, *bathe* and *clothe*, are assumed to be common enough to conduct minimal pair tests.

[2] To prevent speakers from focusing excessively on the target stimuli, words were embedded within meaningful passages rather than isolated carrier sentences. Consequently, the 75 target words were distributed across three separate passages to maintain narrative coherence.

minimal pairs were extracted from the recordings.

3.2 Implementation of Identification Tasks

Listeners were 18 speakers of English (9 men and 9 women), whose L1s include English (n = 5), Chinese (n = 5), Indonesian (n = 3), Telugu (n = 2), Hungarian (n = 1), Korean (n = 1), and Spanish (n = 1). All of them reported that they had spoken English for at least four years. All listeners were 18 years old or older, with the majority in their 20s (n = 12). Eight of the listeners self-reported that they speak Japanese as an additional language, including L1 English listeners (n = 3), Chinese listeners (n = 2), Indonesian listeners (n = 2), and a Korean listener (n = 1).

A minimal pair test was employed as an identification task. The same minimal pairs used to select speech stimuli were employed in the tasks (i.e., 25 pairs per speaker, totaling 75 pairs). In the tasks, they listened to audio files that presented each of the target words extracted from the recordings in a randomized order. They then selected what they believed they heard from three options: the target word, the foil word, or *neither*. If they chose *neither*, they were asked to specify what they believed was uttered.

The data were collected through Google Forms. The online survey consists of two parts: a background questionnaire and identification tasks. The listeners first answered a background questionnaire, in which they reported their age, gender, L1, and whether they speak Japanese (self-report). After completing the background questionnaire, they performed the identification tasks.

3.3 Data Analysis

Intelligibility scores were calculated as the proportion of correct answers (i.e., the number of the target words selected divided by the total number of questions) separately for vowels and consonants. When *neither* was selected, a specified answer was checked manually, and it was counted as correct when a phoneme of interest was correctly understood[3]. Two three-way ANOVAs were computed for vowels and consonants separately, with speaker exposure to English as a within-subject factor and with listener knowledge of Japanese and listener L1 status as between-subject factors. The significance level was set at .05, and all pairwise comparisons were conducted with Tukey-adjusted p-values. As effect sizes, partial eta squared (η_p^2) and Cohen's d were used for ANOVAs and pairwise comparisons, respectively. They were interpreted following the conventions, where η_p^2 values (.01, .06, and .14) correspond to small, medium, and large effects, and d values (0.2, 0.5, and 0.8) correspond to small, medium, and large effects, respectively

[3] For example, one minimal pair for /b/-/p/ was *tab-tap*, where *tab* is the target word. If a listener selected *neither* and answered *tub*, this answer was regarded as correct because the listener correctly understood /b/, or the phoneme of interest in this pair.

Given that the data are proportional, an arcsine transformation was applied to address the assumption of normality and the homogeneity of variance. Following the transformation, Shapiro-Wilk tests on the model residuals verified normality (W = .987, p = .819 for vowels; W = .986, p = .760 for consonants). Additionally, Levene's tests confirmed the homogeneity of variance for the between-subject factors for both vowels (F (3,50) = .316, p = .814) and consonants (F (3,50) = .756, p = .524). Sphericity for the within-subject factor (i.e., speaker exposure to English) was also confirmed via Mauchly's tests (W = .917, p = .569 for vowels; W = .871, p = .409 for consonants). All inferential analyses (i.e., ANOVAs and Tukey-adjusted t-tests) were computed with the arcsine-transformed data, while the descriptive analyses (i.e., means and standard deviations) are reported with the raw data for ease of interpretation.

4 Results and Discussion

The three-way ANOVAs revealed no significant interactions between any factors (p > .05; see Table 2). Therefore, I focus on the main effects for each variable independently.

Table 2

ANOVA Summary Table

	Vowels				Consonants			
Effects	df	*F*-value	*p*-value	η_p^2	df	*F*-value	*p*-value	η_p^2
Between-subject								
L1 status (L1)	1, 14	3.76	.073	.21	1, 14	1.63	.222	.10
Japanese knowledge (JK)	1, 14	4.83	**.045**	.26	1, 14	0.55	.472	.04
L1: JK	1, 14	0.49	.493	.03	1, 14	0.12	.736	.01
Within-subject								
Speaker (SP)	2, 28	3.48	**.045**	.20	2, 28	7.60	**.002**	.35
SP: L1	2, 28	0.70	.505	.05	2, 28	0.14	.867	.01
SP: JK	2, 28	0.93	.405	.06	2, 28	0.83	.447	.06
SP: L1: JK	2, 28	0.85	.440	.06	2, 28	0.60	.556	.04

Note. *P*-values less than .05 are bolded. For the interpretation of an effect size (η_p^2), > .01 is small, > .06 is medium, and > .14 is a large effect.

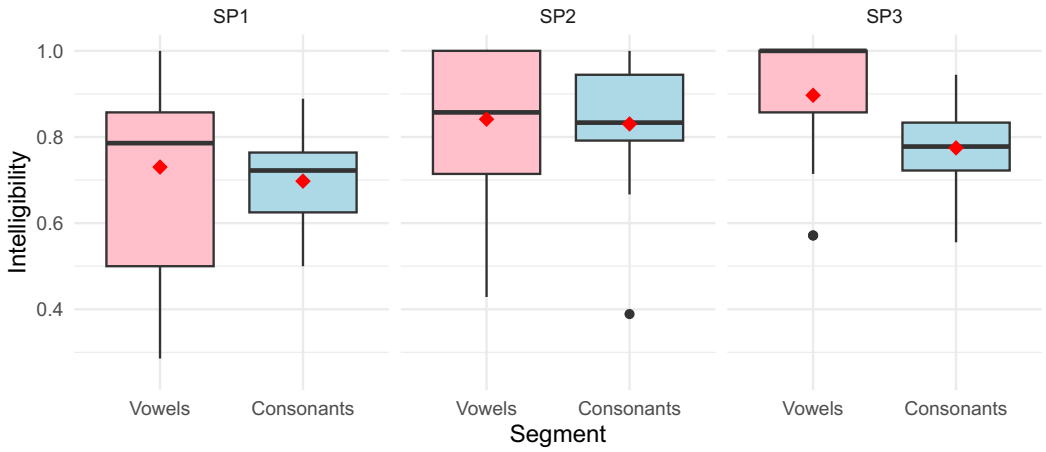
4.1 Speaker Exposure to English

The ANOVAs showed a significant effect of speaker exposure to English on the intelligibility of both vowels and consonants. Post-hoc tests revealed that for vowels, SP3 (M_{SP3} = .897, SD_{SP3} = .145) was significantly more intelligible than SP1 (M_{SP1} = .730, SD_{SP1} = .234; t(28) = 2.84, p = .022, d = 1.27). No significance was

found between SP1 and SP2 (M_{SP2} = .841, SD_{SP2} = .169; $t(28) = 2.09$, p = .110, d = 0.94), and between SP2 and SP3 ($t(28) = 0.75$, p = .738, d = 0.34). For consonants, SP2 (M_{SP2} = .830, SD_{SP2} = .144) was significantly more intelligible than SP1(M_{SP1} = .698, SD_{SP1} = .105; $t(28) = 3.14$, p = .011, d = 1.41). No significance was found between SP1 and SP3 (M_{SP3} =.775, SD_{SP3} = .102; $t(28) = 1.85$, p = .171, d = 0.83), and between SP2 and SP3 ($t(28) = 1.28$, p = .417, d = 0.57). Figure 1 displays the box plots of the intelligibility scores for the three speakers, with red dots indicating the mean scores.

Figure 1

Intelligibility Scores of the Three Speakers



Note: Intelligibility scores were calculated as the proportion of the correct answers to account for the different numbers of minimal pairs for vowels and consonants.

These results, in general, suggest that increased exposure to English has a positive impact on L2 intelligibility. Nonetheless, while the speaker with the least exposure (i.e., SP1) showed the lowest score for vowels and consonants, the most exposed speaker (i.e., SP3) did not consistently show the highest. In fact, no significant difference was observed between SP2 and SP3 for both vowels and consonants. This means that SP2 is as intelligible as SP3 despite the difference in their exposure to English: SP2's intensive exposure to English was limited to inside the university, and SP3 had daily exposure to English. This could be attributed to the ceiling effect; although SP2 had less exposure than SP3, his experience in the English-medium program and the exchange program may have been sufficient for him to reach an intelligibility threshold beyond which additional exposure provides less benefit for intelligibility.

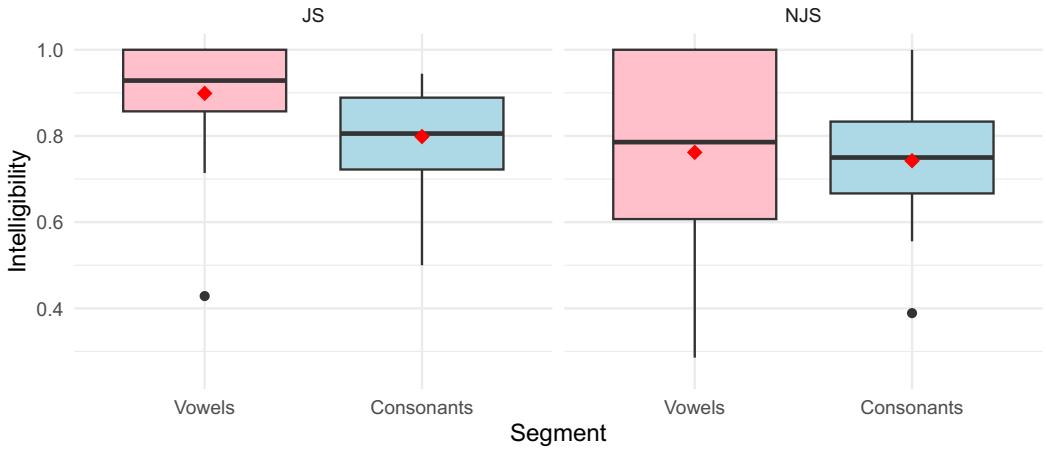
4.2 Listener Knowledge of Japanese

The ANOVA for vowels revealed significance in this factor, whereas the effect for consonants was not significant. Post-hoc test, however, revealed no significance between Japanese-speaking (JS) listeners (M_{JS}

= .899, SD_{JS} = .136) and non-Japanese-speaking (NJS) listeners (M_{NJS} = .762, SD_{NJS} = .217; $t(14) = 1.45$, p = .170, d = 0.53) for vowels. For consonants, JS listeners (M_{NJS} = .799, SD_{NJS} = .111) showed slightly higher scores than NJS listeners (M_{NJS} = .743, SD_{NJS} = .138) without significance. Figure 2 plots the intelligibility scores by JS listeners and NJS listeners.

Figure 2

Intelligibility Scores by Japanese-Speaking (JS) and Non-Japanese-Speaking (NJS) Listeners



These results apparently indicate that the intelligibility of Japanese-accented English is generally comparable between JS and NJS listeners. However, the fact should be noted that the ANOVA showed a significant effect on vowel intelligibility whereas the post-hoc pairwise comparison did not. One possible account for this difference is a lack of statistical power due to the sample size: 18 listeners may not have provided sufficient power to detect significance in a specific pairwise comparison. Nonetheless, given that the ANOVA showed a significant effect and that the ANOVA and the post-hoc test indicated medium to large effects (η_p^2 = .26 and d = 0.53), there might be a trend that knowledge of speakers' L1 (i.e., Japanese) as an additional language provides an advantage in the judgment of vowel intelligibility in speakers' L2 (i.e., English). This trend is specific to vowels, as an ANOVA for consonants revealed no significance or large effect ($F(1,14) = 0.55$, p =.472, η_p^2 = .04). While the trend on the vowel intelligibility could be interpreted as the ISIB, the absence of a significant interaction effect between listener knowledge of Japanese and listener L1 status distinguishes it from the classic mismatched ISIB. In the mismatched ISIB, L2 listeners have an advantage in L2 intelligibility judgment over L1 listeners, regardless of listeners' specific L1 backgrounds (Bent & Bradlow, 2003). In contrast, the current results could suggest that JS listeners have a general advantage over NJS listeners, whether they are L1 listeners or not. This implies that the ISIB might be extended to listeners who acquired speakers' L1 as an additional language.

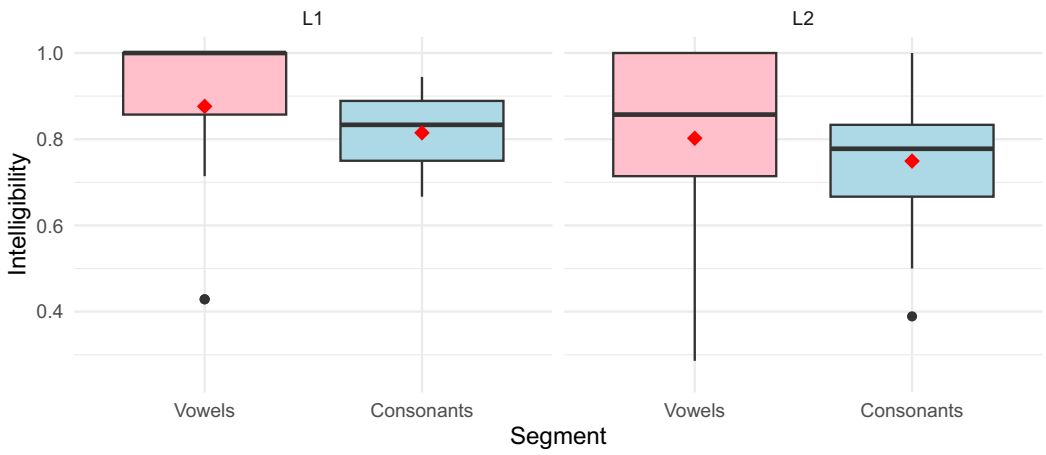
4.3 Listener L1 Status

The ANOVAs revealed no significant effect of listener L1 status for either vowels or consonants. Figure 3 illustrates the box plots of the intelligibility scores by L1 and L2 English listeners. Although the difference did not reach statistical significance, the L1 listener group (M_{L1} = .876, SD_{L1} = .201 for vowels; M_{L1} = .815, SD_{L1} = .086 for consonants) showed slightly higher scores than the L2 listener group (M_{L2} = .802, SD_{L2} = .193 for vowels; M_{L2} = .749, SD_{L2} = .138 for consonants).

These results indicate that L2 listeners are comparable to L1 listeners in the judgement of intelligibility. This contrasts with the findings from Jułkowska and Cebrian (2015), in which the L1 listener group outperformed the other groups. This difference may stem from methodological differences. In their research, intelligibility was operationalized through transcription tasks, and it was measured by the number of words transcribed correctly divided by the total number of words. On the other hand, this study employed minimal pair identification tasks that allowed listeners to choose an answer from options. This likely decreased the rate of inaccurate answers of the L2 listener in this study. It should also be noted that the L2 listeners' proficiency level in English was not controlled in this study. Therefore, it might be possible that the L2 listeners were all proficient enough to blur the distinction between the listener groups.

Figure 3

Intelligibility Scores by L1/L2 English Listeners



5 Conclusion

This study investigated the influence of extralinguistic factors on the intelligibility of Japanese-accented English (JaE). The findings suggest that speaker exposure to English plays a more critical role in judgments of L2 intelligibility than listener knowledge of speakers' L1 and listener L1 status. While the results underscore the multifaceted nature of L2 perception, several limitations must be acknowledged. First, this study had 18 listeners and three speakers. Although the required assumptions for an ANOVA were met, this sample size may

have affected the results of the inferential analyses. Second, because the stimuli were selected based on a general analysis of JaE from the Speech Accent Archive, the individual speakers recorded for this study did not consistently exhibit every target phonemic substitution. In fact, certain targets were correctly identified by all the listeners (e.g., SP2's /b/ and SP3's /oʊ/). Thirdly, this study defined listener knowledge of Japanese as their self-reported ability to speak Japanese, and their proficiency level was not controlled. In addition, L1 listeners of Japanese were not included in the listener group. Collectively, these limitations necessitate cautious interpretation of the potential "extended" ISIB.

Acknowledgements

This project was part of my Bachelor's degree at Waseda University. I am deeply grateful to Victoria Muehleisen for her supportive supervision during my Bachelor's degree and to Mary O'Brien for her informative feedback on improving this manuscript. I would also like to thank two anonymous reviewers for their constructive feedback.

References

Aoyama, K., Hong, L., Flege, J. E., Akahane-Yamada, R., & Yamada, T. (2023). Relationships between acoustic characteristics and intelligibility scores: A reanalysis of Japanese speakers' productions of American English liquids. Language and Speech, 66(4), 1030–1045.

Bent, T., & Bradlow, A. R. (2003). The interlanguage speech intelligibility benefit. The Journal of the Acoustical Society of America, 114(3), 1600–1610. https://doi.org/10.1121/1.1603234

Davies, M. (2008-). Word frequency data. The Corpus of Contemporary American English (COCA). https://www.english-corpora.org/coca/

Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility: Evidence from four L1s. Studies in Second Language Acquisition, 19(1), 1–16.

Gass, S., & Varonis, E. M. (1984). The effect of familiarity on the comprehensibility of nonnative speech. Language Learning, 34(1), 65–87. https://doi.org/10.1111/j.1467-1770.1984.tb00996.x

Huensch, A., & Nagle, C. (2021). The effect of speaker proficiency on intelligibility, comprehensibility, and accentedness in L2 Spanish: A Conceptual Replication and Extension of Munro and Derwing (1995a). Language Learning, 71(3), 626–668. https://doi.org/10.1111/lang.12451

Jenkins, J. (2002). A sociolinguistically based, empirically researched pronunciation syllabus for English as an international language. Applied Linguistics, 23(1), 83–103.

Jułkowska, I. A., & Cebrian, J. (2015). Effects of listener factors and stimulus properties on the intelligibility, comprehensibility, and accentedness of L2 speech. Journal of Second Language Pronunciation, 1(2), 211–237.

Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation Teaching. TESOL Quarterly, 39(3), 369. https://doi.org/10.2307/3588485

Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. Language Learning, 45(1), 73–97.

Munro, M. J., Derwing, T. M., & Morton, S. L. (2006). The mutual intelligibility of L2 speech. Studies in Second Language Acquisition, 28(1), 111–131.

Nagle, C. L., & Huensch, A. (2020). Expanding the scope of L2 intelligibility research. Journal of Second Language Pronunciation, 6(3), 329–351. https://doi.org/10.1075/jslp.20009.nag

Nishi, K. (2002). Perceptual characteristics and intelligibility of Japanese-accented English [PhD dissertation]. In University of South Florida.

Pesantez, A., & Lima, R. J. (2023). Assessing intelligibility of L2 speech by native and non-native listeners: a longitudinal study. Proceedings of the International Congress of Phonetic Sciences, 2482–2486.

Watanabe, Y. (2017). The conflation of /l/ and /r/: New Zealand perceptions of Japanese-accented English. Language Awareness, 26(2), 134–149. https://doi.org/10.1080/09658416.2017.1319849

Weinberger, S. (2015). Speech Accent Archive. George Mason University. http://accent.gmu.edu

Appendix A

Minimal Pairs Used in the Identification Tasks

Contrasts	Target 1	Foil 1	Target 2	Foil 2	Target 3	Foil 3
Vowels						
/i/-/ɪ/*	*seat*	*sit*	*knit*	*neat*	*fit*	*feet*
/ɑ/-/ɔ/	*cards*	*cords*	*part*	*port*	*hard*	*hoard*
/ɔ/-/æ/	*bought*	*bat*	*caught*	*cat*	*sought*	*sat*
/ɔ/-/ɑ/	*bought*	*bot*	*caught*	*cot*	*port*	*part*
/ɔ/-/ʌ/	*bought*	*but*	*caught*	*cut*	*talk*	*tuck*
/eɪ/-/e/	*date*	*debt*	*gate*	*get*	*late*	*let*
/oʊ/-/ɔ/	*boat*	*bought*	*coat*	*caught*	*woke*	*walk*
Consonants						
/b/-/p/	*tribe*	*tripe*	*stable*	*staple*	*tab*	*tap*
/t/-/d/	*bat*	*bad*	*debt*	*dead*	*pot*	*pod*
/t/-/tʃ/	*two*	*chew*	*till*	*chill*	*tech*	*check*
/d/-/t/	*code*	*coat*	*bend*	*bent*	*rude*	*route*
/g/-/k/	*pig*	*pick*	*bag*	*back*	*anger*	*anchor*
/v/-/f/	*leave*	*leaf*	*prove*	*proof*	*live*	*life*
/θ/-/s/	*theme*	*seem*	*thought*	*sought*	*thin*	*sin*
/θ/-/t/	*theme*	*team*	*thought*	*taught*	*thrill*	*trill*
/ð/-/d/	*though*	*dough*	*this*	*diss*	*they*	*day*
/ð/-/s/	*though*	*so*	*this*	*sis*	*they*	*say*
/ð/-/z/	*breathe*	*breeze*	*clothe*	*close*	*bathe*	*bays*
/z/-/θ/	*close*	*cloth*	*phase*	*faith*	*use*	*youth*
/z/-/s/	*seas*	*cease*	*phase*	*face*	*prize*	*price*
/ʃ/-/s/	*dish*	*diss*	*ash*	*ass*	*push*	*puss*
/ʤ/-/ʦ/	*cards*	*carts*	*kids*	*kits*	*sides*	*sites*
/ɹ/-/l/	*brush*	*blush*	*free*	*flee*	*read*	*lead*
/l/-/ɹ/	*blue*	*brew*	*play*	*pray*	*glad*	*grad*
/w/-/v/	*wine*	*vine*	*west*	*vest*	*wet*	*vet*

* Due to a stimulus coding error, these minimal pairs were excluded from the final analysis.

Pronouns in Japanese Junior High School English Education: A Cross-Disciplinary Inquiry into their Usages and Functions Based on the National Curriculum and Standardized Testing

Masayuki Takahashi[1] and Kiyomi Okamoto[2]

[1]Osaka Seikei University, [2]The University of Osaka

takahashimas@osaka-seikei.ac.jp, kiyomiokamoto.cme@osaka-u.ac.jp

Abstract

This study investigates the usages and functions of English pronouns appearing in junior high school materials in Japan, bridging curriculum and empirical perspectives. Drawing on the Course of Study as the national curriculum and the National Assessment of Academic Ability (hereafter, NAAA) as standardized testing data, the study aims to clarify how pronouns are treated pedagogically and functionally. Two approaches were adopted: (a) quantitative analysis of pronoun frequency and grammatical category based on the Revised List of Grammar Items (Takahashi & Okamoto, 2025), and (b) interpretation of how such results inform curriculum alignment and grammar instruction. Combining the curriculum's institutional framework with authentic test data, this study represents the first attempt to depict the actual distribution and discourse functions of pronouns in officially sanctioned materials. Findings show there are 34 distinct pronouns, including five that are not listed in the Course of Study (both, each other, everyone, something, everything). Comparison with the British National Corpus revealed that the pronouns "I," "you," "it," "we," and "they" dominate in the NAAA, highlighting a partial misalignment between prescriptive curriculum items and actual usage patterns.

Keywords

pronouns, grammar instruction, Course of Study, National Assessment of Academic Ability, Revised List of Grammar Items

1 Purpose and Significance of the Study

The purpose of this study is to clarify the usage and functions of English pronouns appearing in texts used at the junior high school level in Japan. For this purpose, the study analyzes English sentences drawn from two primary sources: (a) instructional content based on the Course of Study (Ministry of Education, Culture, Sports, Science and Technology [MEXT], 2018a; hereafter, CoS), the national curriculum, and (b) English passages used in standardized tests, the National Assessment of Academic Ability (hereafter, NAAA).

This study has two methodological pillars. The first involves analyzing the frequency and usage of pronouns in the texts based on the Revised List of Grammar Items[4]. The second considers how the results of this analysis can be applied to enhance grammar instruction, with reference to the CoS.

By combining the institutional (CoS) and empirical (NAAA) perspectives, this study uniquely aims to clarify the reality of pronoun usage in English education in Japanese junior high schools. It also aims to identify what is taught as grammatical categories of pronouns, examines how these relate to actual use, and highlights gaps or consistencies that may inform teaching practices and curriculum improvement.

2 Background

2.1 The Course of Study and Grammar

The CoS establishes national standards to maintain consistent educational quality from kindergarten through high school. In English education for junior high schools, the CoS presents language materials classified as "Items related to the characteristics and rules of English," composed of four domains: "Sounds," "Symbols," "Words, collocations and idiomatic expressions," and "Sentences, sentence structures and grammar items." Within this framework, the CoS systematically specifies the grammatical categories and instructional items for English grammar in junior high schools.

2.2 Developing a Comprehensive Grammar List

One of the motivations for developing the List of Grammar Items[4] as a comprehensive grammar list was the lack of any comprehensive catalogue of grammatical items in Japan, as well as the absence of prior large-scale studies that presented a systematic inventory of such items. Although several smaller-scale studies are listed below, they cannot be regarded as a true comprehensive overview.

Nonetheless, it should be clearly stated that these earlier works served as an important point of departure for the creation of the List of Grammar Items. Chujo *et al.* (2012) examined remedial learners' grammatical mastery. Ushie (2013) identified problems in junior high school textbook grammar descriptions, focusing on relative and contact clauses, the there-construction, tense, comparison, contractions, fixed expressions, and ellipsis. Yasugi (2016) specified a wide range of frequently problematic items such as the be-verb, do/does-questions, the progressive, comparison, infinitives, gerunds, the passive, the present perfect, post-modification, relative pronouns, wh-questions, SVOO/SVOC, and indirect questions. Tanaka (2017) demonstrated that 96.3% of TOEIC items can be solved using junior high school grammar and identified key high-frequency categories including nouns, adverbs, adjectives, prepositions, comparison, tense, infinitives, and relative

[4] Unless otherwise noted, all references to the Revised List of Grammar Items and the List of Grammar Items in this paper refer to the version presented in Takahashi and Okamoto (2025) and Takahashi and Okamoto (2024), respectively.

clauses. Honda and Shimura (2017) compared elementary and junior high school grammar lists, presenting major categories (e.g., pronouns, nouns, verbs, auxiliaries, tense, interrogatives, comparison) and noting that their 162 subcategories are excessively fine-grained.

2.3 Systematizing Grammar Items

The *List of Grammar Items* and the *Revised List of Grammar Items* were developed within a KAKENHI-funded research project that aimed to construct a model for teaching English designed to increase learners' noticing of grammar through the LOGIC (List of Grammar Items and Corpus) system. At the core of this educational model lies the LOGIC system, which provides examples of English expressions relevant to pedagogical grammar in junior high schools. This line of research aims to contribute to English education by constructing a model applicable both to classroom instruction and to learners' independent study.

The Revised List of Grammar Items was designed to clarify school grammar in Japanese junior high schools and to present a grammatical framework applicable to teaching practice. It organizes 477 grammar items into 22 grammar categories. The revision of the list was motivated by empirical findings that emerged as the List of Grammar Items evolved. After detailed examination of its sources—the CoS and NAAA texts—revisions were made to address missing or insufficient items. Consequently, two grammar items concerning pronouns and three concerning prepositions were added during the revision process of the Revised List of Grammar Items.

2.4 The Position of Pronouns in English Education

Within the CoS, 19 grammatical frameworks are listed under "Sentences, sentence structures and grammar items," among which pronouns are explicitly included. The CoS's classification of pronouns encompasses personal, demonstrative, interrogative, and numerical pronouns. Although relative pronouns are also included in the CoS, this study focuses on the four aforementioned categories.

2.5 Pronouns in Takahashi & Okamoto (2025)

In the Revised List of Grammar Items, 27 grammar items concern pronouns (see Table 1). Traditionally, as noted by Yagi (2021) and O (2017), pronouns are classified into personal, demonstrative, interrogative, indefinite, and relative categories. While the CoS identifies only four of these (personal, demonstrative, interrogative, and numerical), the Revised List offers a more comprehensive coverage to strengthen instructional clarity.

Table 1

Pronouns in the Revised List of Grammar Items

Grammar categories	Grammar items	
D001	Referring to specific people or things	Subjective case
D002		Possessive/Objective case
D003	Referring to general people	Subjective case
D004		Possessive/Objective case
D005	Referring to specific people or things	Possessive pronouns
D006	Referring to general people	Possessive pronouns
D007	Expressions with "it" as the subject	(e.g., it is time...)
D008	Reflexive pronouns	Referring to people (specific or general)
D009		Referring to "it" or things
D101		Special expressions: e.g., help oneself
D102		Idiomatic expressions: e.g., by oneself
D010	Demonstrative pronouns	that/these, etc.
D011	Interrogative pronouns	who/whose/whom
D012		which/what
D013	Indefinite pronouns	Usage: one
D014		Usage: other/others/another
D015		Usage: the other/the others
D016	Quantifying pronouns	Usage: few/a few/little/a little/many/much
D017		Usage: some/any
D018	Usage of pronouns	both (including we both/both of us)
D019		either
D020		neither
D021		all (e.g., we all know.../we are all trying...)
D022		none (complete negation)
D023		each
D024	Compound-form indefinite pronouns	someone/everything/something, etc.
D025	Others	each other, etc.

3 Methods

This study analyzed English tests used in the NAAA to investigate the frequency, types, grammatical functions, and contextual uses of pronouns.

3.1 The National Assessment of Academic Ability (NAAA)

The NAAA is a nationwide standardized test conducted by MEXT with the objective of “ensuring equal educational opportunities and improving the level of compulsory education” (MEXT, 2018b, 1. Purpose of the Survey). The NAAA is grounded in the CoS, and the full texts and correct answer examples used in the tests are publicly available.

The assessment is conducted every April for sixth-grade elementary and ninth-grade junior high school students. While Japanese and mathematics are tested annually, English and science are tested every three years. The English test targets third-year junior high school students (Grade 9) who have studied English formally for four years.

The English component has been implemented three times: the preliminary test (hereafter, Test 0) in 2018, the first test (hereafter, Test 1) in 2019, and the second test (hereafter, Test 2) in 2023. The four skill domains—listening, reading, writing, and speaking—cover learning items up to the preceding grade. Across all versions, the dataset includes 523 sentences and 3,755 words (see Table 2).

Table 2

Number of Sentences and Words in the NAAA English Texts

		Listening	Reading	Writing	Speaking	Total
Sentences	Test 0	33	69	23	25	150
	Test 1	41	92	24	20	177
	Test 2	56	89	16	35	196
Words	Test 0	263	525	123	147	1058
	Test 1	282	666	131	116	1195
	Test 2	435	703	103	261	1502

3.2 Analytical Procedure

Using the Revised List of Grammar Items as the analytical framework, all pronouns in the NAAA texts were identified, categorized, and analyzed for their grammatical and contextual functions. Leech *et al.* (2001) was employed as a comparative corpus resource.

4 Results and Discussion

4.1 Pronouns Identified

A total of 34 distinct pronouns described below were identified in the NAAA English texts. Five of the distinct pronouns—both, each other, everyone, something, and everything—were not listed in the CoS, a noteworthy finding that demonstrates lexical expansion beyond the prescribed curriculum.

- Personal pronouns: I, my, me, mine, you, your, he, his, him, she, her, it, its, we, our, us, they, their, them
- Demonstrative pronouns: this, that, these
- Interrogative pronouns: who, which, what
- Numerical/quantifying pronouns: none, one, some, all
- Non-CoS pronouns: both, each other, everyone, something, everything

4.2 Frequently Used Pronouns: I, you, it, we, they

The pronouns “I,” “you,” “it,” “we,” and “they” were the most frequently occurring forms. The following are excerpts of its occurrences. The notation in brackets [ ] indicates the grammar category in the Revised List of Grammar Items, and the corresponding description on the right shows the specific grammar item. In addition,

the label at the end of each excerpt identifies the NAAA test year (e.g., Test 0 for 2018) and the skill area (Listening, Reading, Writing, or Speaking). For instance, "(Test 0, Listening)" indicates that the excerpt was taken from the listening section of the test conducted in 2018. These are the findings:

[D001] Referring to specific people or things: Subjective case
I want to sing with Mike. (Test 0, Listening)
It's very popular. (Test 2, Listening)

[D002] Referring to specific people or things: Possessive/Objective case
Now, I'm going to tell you about its history. (Test 0, Reading)
How about this bag with some stars on it? (Test 2, Listening)

[D003] Referring to general people: Subjective case
For example, in the U.S., they have one-dollar shops. (Test 1, Reading)
We should not buy too much food. (Test 1, Reading)

These excerpts show that high-frequency pronouns are core personal pronouns, consistently taught from the earliest stages of instruction. Their prominence emphasizes the functional importance of identifying speakers, listeners, and topics, which are key elements in communicative English education.

4.3 Comparison with the British National Corpus

This study employed the British National Corpus (BNC) in Leech *et al.* (2001) as a reference corpus. The list of pronouns appearing in the BNC is provided in List 5.5: Frequency list of pronouns (not lemmatized) in Chapter 5, Rank Frequency Lists of Words within Word Classes (Parts of Speech) in the Whole Corpus. Table 3 below presents the top 20 pronouns found in the BNC, showing their corresponding rank (Test rank) and frequency (Test freq.) in the NAAA.

While the pronouns "he", "she", and "they" are among the most frequent in the BNC (ranks 4th, 8th, and 7th, respectively), they appear only occasionally in the NAAA. In contrast, "I", "you", and "we" show relatively higher frequencies in the NAAA (ranks 1st, 3rd, and 11th, respectively). This suggests that the NAAA emphasizes familiar, learner-centered pronouns associated with direct communication, rather than narrative or descriptive forms.

Table 3

Top 20 Pronouns in the British National Corpus and their Rank/Frequency in the NAAA

BNC rank	Pronouns	Test rank	Test freq.	BNC rank	Pronouns	Test rank	Test freq.
1	it	3	57	11	we	4	40
2	I	1	92	12	her	29	1
3	you	2	70	13	their	17	6
4	he	11	10	14	what	9	13
5	this	17	6	15	all	22	3
6	his	22	3	16	who	22	3
7	they	5	34	17	them	12	9
8	she	13	7	18	some	25	2
9	that	19	5	19	him	13	7
10	which	25	2	20	its	25	2

4.4 Usage of Selected Pronouns

4.4.1 Quantifying Pronouns: some/any

The pronouns “some” and “any,” often introduced as determiners (e.g., some pictures), also occur as independent pronouns. While “some” appears in the CoS, “any” does not. The CoS excerpts include:

> If you want water, I will give you some. (p. 42)
>
> We have other bigger ones. Shall I show you some of them? (p. 71)

The NAAA excerpts include:

> [D017] Quantifying pronouns Usage: some/any
>
> I want to try some! (Test 1, Listening)
>
> I was surprised that some of these shops sell medicine. (Test 1, Reading)

These indicate that both “some” and “any” deserve deliberate instructional attention, as they are conceptually simple but contextually sensitive items.

4.4.2 Indefinite Pronouns: one

This section focuses on the word “one”, which can function as a noun, an adjective, or a pronoun. The pronoun “one” is used as a substitute for a previously mentioned countable noun or noun phrase. It typically co-occurs with determiners such as adjectives or articles:

[D013] Indefinite pronouns/Usage: one

Which one? (Test 0, Listening)

The one with three birds and some English words on it. (Test 0, Listening)

Some students use an electronic one when they don't know a word. (Test 0, Reading)

The chimpanzees got food when they pointed to the stronger one. (Test 1, Reading)

I'll take the bigger one. (Test 2, Listening)

Which one should I buy for him, a picture book, animal cookies or a T-shirt? (Test 2, Speaking)

The following are example sentences listed in the CoS.

"A: Which bus goes to the city library? B: Take this one." (p. 38)

This question is more difficult than that one. (p. 48)

"A: We have other bigger ones. Shall I show you some of them? B: Thank you. Oh, this one seems better. How much is it?" (p. 71)

Intentional instruction targeting the contrast between the singular "one" and the plural "ones" can broaden learners' expressive options and is likely to support the development of the communicative competence that the CoS seeks to foster.

4.4.3 both (including we both / both of us)

The pronoun "both," which is not listed in the CoS, appeared in the NAAA. This was considered a noteworthy finding. The following is the excerpt.

[D018] both (including we both/both of us)

We both love sports, music, and cooking, right? (Test 2, Reading)

The sentence above is included in the CoS under "(ii) Items related to language functions," where it appears in the section stating, "Pick up language use situations and language functions mainly from the following items when engaging in language activities." (p. 69). In addition, "b. Examples of language functions" are organized into five categories: "(a) Facilitating communication," "(b) Communicating feelings," "(c) Communicating facts and information," "(d) Communicating thoughts and intentions," and "(e) Encouraging someone to do something." The fact that the above example is treated within this section of the CoS suggests that it is intended as an essential expression for fostering communicative competence.

4.4.4 all (e.g., we all.../we are all...)

With regard to “all” as a pronoun, this study focuses on two specific usages: cases in which “all of” is followed by plural noun phrases or noun phrases limited to definite articles, demonstratives, or possessive forms of personal pronouns as the object of “of,” and those in which “all” is used in apposition to the subject or object. Therefore, instances where “all” is accompanied by a relative clause, or used independently to mean “all people” (plural) or “everything” (singular), are not examined here. The following are excerpts from the NAAA.

[D021] all (e.g., we all.../we are all...)
We all said, “We cannot do something big, but doing something to save water is important for even junior high school students.” (Test 2, Listening)
We are all trying to do something to save water. (Test 2, Listening)
I hope all of you visit our creative library in the near future. (Test 2, Reading)

Incidentally, there is the expression “first of all.” In the CoS, the example “First of all, let me talk about the outline of the story.” (p. 76) is provided. Syntactically, the word “all” functions as the object of the preposition “of” and therefore can be classified as either a noun or a pronoun. It should be noted, however, that this study was unable to determine conclusively whether “all” in this expression functions as a noun or a pronoun.

4.4.5 none (complete negation)

A pronoun “none” corresponds to the pronominal use of “no.” In this study, attention is paid to cases where “none” is used synonymously with “no one” or “nobody” to mean “no person,” as well as to cases where “none of” is followed by plural or uncountable noun phrases introduced by determiners such as the, one’s, or these (e.g., none of the students, none of one’s work, none of these ideas) to express meanings like “nothing at all” or “no one at all.” The following is the excerpt.

[D022] none (complete negation)
In *janken*, none is the strongest among rock, scissors, and paper. (Test 1, Reading)

The pronoun “none” is not listed anywhere in the CoS. This suggests that, although “none” appeared in the NAAA, teachers need not overreact to its presence. Instead, they should focus on fostering students’ communicative competence through instruction aligned with the CoS.

5 General Discussion

This study analyzed pronoun use in Japanese junior high school English education through both the institutional framework (CoS) and empirical evidence (NAAA). The Revised List of Grammar Items proved to be an effective analytical tool for linking systematic grammar classification with textual analysis.

Several findings stand out. First, pronouns not explicitly listed in the CoS (both, each other, everyone, something, everything) were found in standardized tests. This indicates a degree of pragmatic diversity beyond what the CoS prescribes, suggesting potential gaps between curriculum design and real language use.

Second, pronoun use reflects the development of discourse-functional competence, not merely structural knowledge. Pronouns serve as markers of cohesion, perspective, and focus in discourse, thus representing an interface between grammar and pragmatics.

These findings point to the need to redefine school grammar as pedagogical grammar—a grammar that integrates cognitive, functional, and discourse-based dimensions of language learning. This study also demonstrates that systematic lists such as the Revised List of Grammar Items help clarify the relationship between the CoS's instructional goals and the linguistic performance measured by standardized tests. This connection provides a new theoretical and empirical foundation for grammar pedagogy.

6 Conclusions

This study uniquely bridges the institutional documents (CoS) and national empirical evidence (NAAA) to depict the usage and functions of pronouns in Japanese junior high school English education. By applying the Revised List of Grammar Items, the study empirically identified the relationship between grammar instruction and actual usage, a link rarely addressed in traditional grammar studies.

Three pedagogical implications emerge:

1. For teachers: While adhering to the CoS, they should respond flexibly to non-listed items appearing in authentic texts, thereby fostering learners' grammatical competence in use.
2. For material developers: Pronoun introduction and recycling should be sequenced according to learners' developmental stages, integrating form, meaning, and discourse function.
3. For researchers: The Revised List of Grammar Items should be employed as a foundation for verifying consistency between actual classroom language use and curriculum expectations, leading toward evidence-based grammar education models.

Future research should quantitatively compare the distribution of pronouns in the NAAA with those in MEXT-approved textbooks to calculate an alignment index. Moreover, discourse-analytic studies should

investigate how pronouns contribute to task performance and meaning construction, further bridging grammar research and pedagogical practice.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Number 23K00785.

References

Chujo, K., Yokota, K., Hasegawa, S., & Nishigaki, C. (2012). Rimediaru gakushu-sha no eigo jukutatsu-do to eigo bunpo jukutatsu-do chosa [Identifying the general English proficiency and distinct grammar proficiency of remedial learners]. *Journal of the College of Industrial Technology, Nihon University B*, *45*, 43-54.

Honda, N., & Shimura, A. (2017). Shogakko gaikokugo katsudo kyozai to chugakko eigo kyokasho ni okeru bunpo komoku no hikaku [A comparison of grammar in elementary school and junior high school English textbooks]. *Journal of Hokkaido University of Education* (*Education)*, *68*(1), 119-135.

Leech, G., Rayson, P., & Wilson, A. (2001). *Word frequencies in written and spoken English: Based on the British National Corpus*. Pearson.

Ministry of Education, Culture, Sports, Science and Technology. (2018a). *Chugakko gakushu shido yoryo (Heisei 29-nen kokuji) kaisetsu: Gaikokugo-hen* [Commentary on the Course of Study for lower secondary schools (2017 Notification): Foreign languages]. Kairyudo-shuppan.

Ministry of Education, Culture, Sports, Science, and Technology. (2018b). *Heisei 31 nendo zenkoku gakuryoku gakushu jokyo chosa ni kansuru jisshi yoryo* [Implementation guidelines for the FY2019 National Assessment of Academic Ability]. https://www.mext.go.jp/a_menu/shotou/gakuryoku-chousa/zenkoku/1411767.htm

O, T., Kawasaki, Y., Kubota, H., Takada, A., Takahashi, K., Tsuchiya, M., Fisher, G., & Yamada, H. (2017). *Sogo eigo Evergreen* [Comprehensive English grammar: Evergreen]. Iizuna-shoten.

Takahashi, M., & Okamoto, K. (2024). Chugakko gakushu shido yoryo ni motozuku “Bunpo komoku ichiran hyo” no kaihatsu [Development of a “List of Grammar Items” based on the Course of Study for junior high school. *Osaka Seikei Educational Research*, *4*, 50-61.

Takahashi, M., & Okamoto, K. (2025). “Kaiteiban bunpo komoku ichiran hyo” no kaihatsu: “Bunpo komoku ichiran hyo” kara “Kaitei-ban bunpo komoku ichiran hyo” e [Development of the Revised List of Grammatical Items: From the original to the revised edition]. *Osaka Seikei University Bulletin* (*Department of Education*), *11*, 107-118.

Tanaka, M. (2017). TOEIC tesuto kara kangaeru chugaku eigo bunpo no juyo-sei to shido naiyo [The significance of junior high school English grammar and the contents of teaching from a viewpoint of the TOEIC Test]. *Nagasaki International University Review*, *17*, 13-21.

Ushie, K. (2013). Chugakko eigo kyokasho ni okeru bunpo kijutsu to goi donyu no mondai-ten: Sunshine English Course no baai [Problems of grammatical descriptions and vocabulary in junior high school English textbooks:

The case of Sunshine English Course]. *Journal of Saitama University*, *Faculty of Education*, *62*(1), 175-190.

Yagi, K. (2021). *Gendai koto eibunpo*: *Gakushu bunpo kara kagaku bunpo e* [An advanced grammar of present-day English: From pedagogical to scientific]. Kaitaku-sha.

Yasugi, S. (2016). Eigo gakushu-sha no bunpo-jo no tsumazuki o herasu tame no teian [A suggestion for reducing EFL learners' grammar errors: From the analysis of the tests of the key sentences of the junior high school English textbooks]. *CASELE Research Bulletin*, *46*, 79-87.

Automated Writing Comparison: The Case Study of English for Law

Ho Thi Phuong Nam1

National Taiwan Normal University

813210101@ntnu.edu.tw, phuongnam1412@ueh.edu.vn

Abstract

There has been a switch from lesson content mastering to skills developing as the movement of compe- tency in the 21st century. To respond to this trend in practical requirements, University of Economics, Ho Chi Minh City has focused on providing English for Specific Purposes courses where students can learn skills relating to their study fields. Various prior studies have been conducted for seeking ways to assess and evaluate English written compositions. Building and validating appropriate scoring scales have yielded positive results in reducing human raters' burden, especially in ESP writing assessment in- cluding many unique writing tasks. Therefore, ESP teachers can have more chances to give feedback and design remedial lessons accordingly for their students. With Artificial Intelligence, teachers are using numerous tools as personal assistants for designing their own scoring scales. This study used Claude 3.5 Sonnet for creating Python tools comparing 224 Law students' compositions. This tool was validated by comparing its automated scores with human ratings and produced a positive correlation. This tool can reflect human ratings and have the potency to reduce teacher scoring workload with some related pedagogical implications.

Key word

Automated scoring scale, Writing comparison, Automated ratings, Human ratings, ESP

1 Introduction

There has been a switch from the traditional concept of 'to master lesson content' to 'skills develop- ment' as the 'competency movement' of the 21st century (Nunan, 2022). To respond to the trend in teaching–learning activities and the practical requirements, various universities in Vietnam are planning to strengthen skills for their learners in specific disciplines. This leads to the focus on designing and providing English for Specific Purposes (ESP) courses where students expect to learn knowledge and skills relating to their fields of studying.

Therefore, University of Economics, Ho Chi Minh City (UEH) is on the way to reinforce its ESP teaching and learning, as it is aware that this is the right time for ESP, and General English is now the responsibility of each student. In the year of 2023, UEH made a corporation between its School of Foreign Languages (SFL)

and six other Schools for this purpose entailing the establishment of four ESP courses. An abundance of prior studies has been conducted for seeking ways to assess and evaluate English written production. Besides building and validating appropriate scoring scales for writing assessment, error analysis research needs to be increasingly focused to offer more additional insights into ways to achieve effectiveness in English writing (Hughes & Hughes, 2020; Zafar, 2016). Numerous previous research has paid great attention to analyze errors in different English language teaching and learning settings; however, there are a few studies which delve into the investigation of identifying the frequency of errors as well as their effect on ESP written communication. This research direction can yield findings which act as a comprehensive guide for ESP classroom teachers to design more remedial lessons for helping their students.

2 Setting of the Study

The ESP courses established by SFL and six other Schools at UEH encompassing four main majors including English for Business, English for Economics, English for Technology, Communication and Design, and English for Law.

Annually, UEH uses results of some standardized English proficiency tests like TOEIC (Test of En- glish for International Communication), TOEFL iBT (Internet-based-test), IELTS (International English Language Testing System) and VPET (Versant Professional English) for admission. These results are evidence for full-time students to skip the first General English module; however, in order to provide students with pertinent knowledge and skills in their fields of study, they are going to do two compulsory ESP courses.

UEH belongs to the business and economics discipline group of universities in Southern Vietnam; as a result, the vast majority of students are attending business and economics academic clusters followed by the technology, communication and design cluster, and the smallest amount of students are joining the law cluster (University of Economics, Ho Chi Minh City, 2024).

The writing section of English for Law modules consists of two main tasks focusing on not only drafting some legal documents like client letters, witness statements, contract claims, or employment tribunal claims but also filling in legal pleadings or legal forms. Writing is a complicated task, especially in the setting of legal writing, it can become arduous since Law students have to perform many duties simultaneously which are using accurate information of legal cases, and grammatically correct structures in appropriate legal forms.

To shed light on helping Law students at UEH better their writing outcomes, when drafting some specific legal documents, templates will be provided. This aligns with controlled composition attributes stating that students need to be given necessary information and forms of the sentences they will use; ultimately, their final products will be similar to the legal forms required (Raimes, 1983). In this study, the legal writing task of drafting a witness statement, which contains the description of events that happened at the crime scene, will be utilized for error analysis.

3 Literature Review

3.1 Using Automated Tailored Scoring Tools

Numerous research has been carried out to prove that assessment is playing an important role in 'shaping students learning,' and thus there is a need for aligning assessment practices with what the classroom pedagogical and intended outcomes are (Chamberlain et al., 2004). According to the needs analysis made before establishing two English for Law modules, students are expected to be familiar with a number of common legal forms with highly accurate information and appropriate forms; as a result, writing activities and assessment must be effective enough for this purpose, then it is not only positively affecting teaching but also learning.

There have been various prior studies conducted for seeking ways to assess and evaluate English written compositions. Among these, with the advent of Artificial Intelligence (AI), building and validating appropriate automated scoring scales have yielded positive results in reducing human raters' burden with great reliability (Ramineni, 2013; Ramineni et al., 2013). Together with the aforementioned idea, many scientists are also calling for further research on these tools as the AI technology is viable and rapidly growing; therefore, the making of a valid and reliable assessment instrument is a must (Ardhy, 2023; Drid, 2018). Considering the legal writing context, where there are many unique writing tasks involved, this renders a fruitful research theme since there are merely a few research attempting to create a specific scoring tool for legal written compositions.

3.2 Error Analysis (EA) in Scoring Legal Written Compositions

Based on the discussion of Corder (1967), error analysis (EA) has been widely used in writing assessment because of at least threefold benefits. EA is rich in data informing learning progress for teachers, how the language is obtained by ESP students for researchers, especially the learning input itself for students (Ulyani, 2024). An abundance of previous studies has paid great attention to detecting grammatical errors and providing explanations for these occurrences accordingly, but they have not recommended related pedagogical implications (Ulyani, 2024).

Furthermore, EA can assist ESP teaching and learning process by offering guidelines on how these errors can be corrected and thus can prevent recurring in the future (Tas´ıc, 2023). Erdogan (2005) introduced the term of 'Applied EA', as compared to the 'Theoretical EA', which delved into the 'remedial courses' and developing appropriate materials and teaching practices based on the findings of errors (Hussain, 2024).

In the context of legal writing assessment, EA can be the combination of frequency of errors and their effects on communication (Hughes & Hughes, 2020). With the effect on communication, legal writing assessment might go beyond the ordinary mechanical aspects and focus more on communicative dimen- sions, which means it cares about clarity, cohesion, accuracy in a specific context (Ardhy, 2023). To serve the purpose of achieving effective legal English writing, analysts need to successfully deal with the most prominent errors

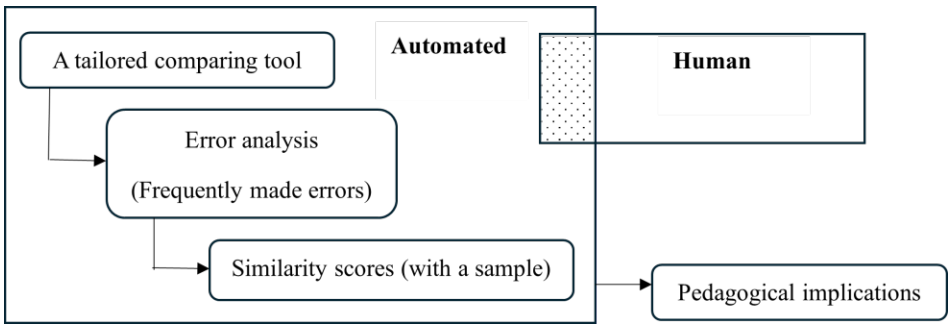
made by students (Zafar, 2016), and by working on written composition's data, EA is considered a practical diagnostic tool for informing instruction and learning (Xie, 2019).

4 Conceptual Framework of the Study

Based on the review of existing literature and research gaps, a conceptual framework (see Figure 1) has been developed in order to highlight the areas that need investigation. In this study, an automated comparing tool was created, and this tool compared students' legal written compositions with a sample text to provide similarity scores. For the sake of validating this automated comparing tool, a correlation statistical test will be conducted to check whether the automated tool can reflect the patterns of human ratings. Ultimately, some pertinent pedagogical implications were discussed accordingly.

Figure 1

Conceptual framework



5 Research Questions

To serve the purpose of finding out the most frequently made errors in students' legal writing production and checking the validity of the newly created comparing tool, two research questions are raised:

1. Which are the types of errors English for Law students frequently make in their written compositions?
2. To what extent does the automated comparing tool tailored for legal writing reflect the human rating patterns?

6 Methodology

6.1 Materials

The materials for this study comprised (1) 224 written compositions drafting a witness statement of students from five English for Law 1 classes, (2) a sample text, (3) the Claude 3.5 Sonnet AI Chatbot, (4) two Python

3.13.3 programs for text comparison and correlation analysis, and (5) human ratings. The students were all in their first academic year, and some of them had just finished their first General English module at UEH. The students who did not attend this first General English module were holding some standardized English proficiency tests; therefore, they were allowed to skip this General English module and continue to take the English for Law 1 module based on the school policy. The human ratings were the scores of these 224 written compositions from three teachers of SFL. In this group of raters, one teacher was a teacher of English for Law and the other two teachers were from other ESP courses.

6.2 Design

This study used an AI Chatbot – Claude 3.5 Sonnet for writing a Python code to create a tailored com- paring tool. The tool's aim was simply at comparing 224 students' legal written compositions with a given sample text and finally reporting on similarity score of each students' composition. This design was aligned with the existing literature, where many educators or teachers have made use of AI tools for evaluating students production, for learning programming languages, and for analyzing data, especially the two latter usages had limited the research ability of some scholars (Chinedu & Ade- Ibijola, 2021; Davar et al., 2025). Finally, this newly created comparing tool was validated by checking its correlation with the human ratings.

6.3 Instruments

In order to run the automated comparing tool properly, besides the collection of 224 students' legal written compositions, a sample text needed to be given as substantial input (see Figure 2). As the writing task in this study required, students had to write a composition with accurate information and use a witness statement form; for this reason, a template was also provided (see Figure 3 and Figure 4).

Figure 2

Sample text

I, John Doe of 303 Maple Avenue, Springfield, am a witness in this claim. The facts in this statement come from personal knowledge or as the case may be.

On 5 March, 2024 at approximately 3:45 PM, I was jogging through the park near Maple Avenue when I noticed a disturbance. I stopped to look and witnessed two young adults shouting at each other near the swings. Suddenly, one of them picked up a stick and threw it in the direction of the other. I saw Mr. Smith, a 65-year-old man, get hit on the leg by the stick.

I immediately called for help and went over to assist Mr. Doe along with other bystanders. We didn't leave until the paramedics/emergency services arrived.

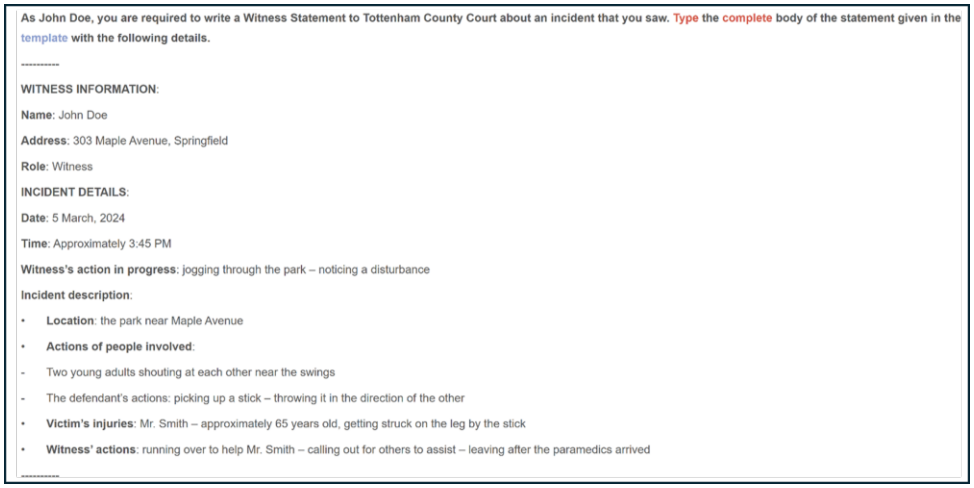
I believe that the facts stated in this Witness Statement are true. I understand that proceedings for contempt of court may be brought against anyone who makes, or causes to be made, a false statement in a document verified by the statement of truth without an honest belief in its truth.

Signed: John Doe

Dated: 5 March, 2024

Figure 3

Information for writing



As John Doe, you are required to write a Witness Statement to Tottenham County Court about an incident that you saw. Type the complete body of the statement given in the template with the following details.

WITNESS INFORMATION:

Name: John Doe

Address: 303 Maple Avenue, Springfield

Role: Witness

INCIDENT DETAILS:

Date: 5 March, 2024

Time: Approximately 3:45 PM

Witness's action in progress: jogging through the park – noticing a disturbance

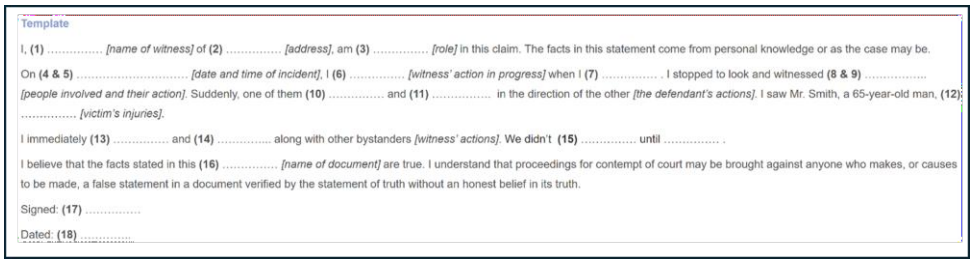
Incident description:

- **Location**: the park near Maple Avenue
- **Actions of people involved**:
- Two young adults shouting at each other near the swings
- The defendant's actions: picking up a stick – throwing it in the direction of the other
- **Victim's injuries**: Mr. Smith – approximately 65 years old, getting struck on the leg by the stick
- **Witness' actions**: running over to help Mr. Smith – calling out for others to assist – leaving after the paramedics arrived

---

Figure 4

Template for writing



Template

I, **(1)** ..... *[name of witness]* of **(2)** ..... *[address]*, am **(3)** ..... *[role]* in this claim. The facts in this statement come from personal knowledge or as the case may be.

On **(4 & 5)** ..... *[date and time of incident]*, I **(6)** ..... *[witness' action in progress]* when I **(7)** ..... I stopped to look and witnessed **(8 & 9)** ..... *[people involved and their action]*. Suddenly, one of them **(10)** ..... and **(11)** ..... in the direction of the other *[the defendant's actions]*. I saw Mr. Smith, a 65-year-old man, **(12)** ..... *[victim's injuries]*.

I immediately **(13)** ..... and **(14)** ..... along with other bystanders *[witness' actions]*. We didn't **(15)** ..... until ..... .

I believe that the facts stated in this **(16)** ..... *[name of document]* are true. I understand that proceedings for contempt of court may be brought against anyone who makes, or causes to be made, a false statement in a document verified by the statement of truth without an honest belief in its truth.

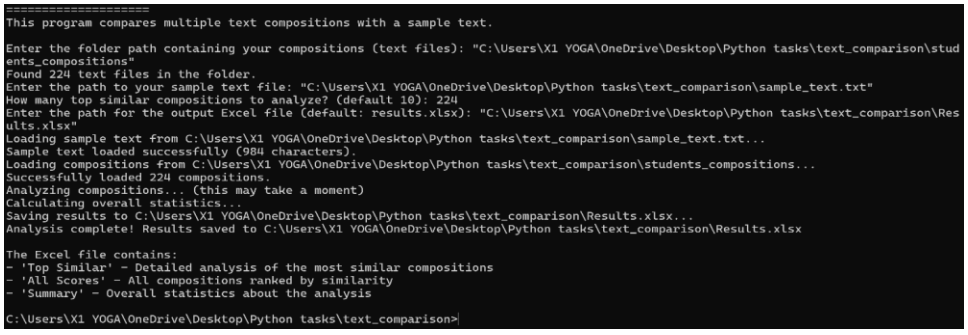
Signed: **(17)** .....

Dated: **(18)** .....

With the support of Claude AI Chatbot, two Python programs for text comparison and correlation analysis were created (see Figure 5 and Figure 6).

Figure 5

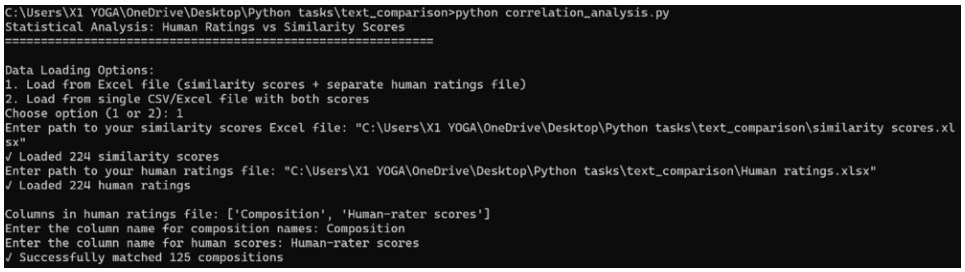
Python program for text comparison



```
==================
This program compares multiple text compositions with a sample text.

Enter the folder path containing your compositions (text files): "C:\Users\X1 YOGA\OneDrive\Desktop\Python tasks\text_comparison\stud
ents_compositions"
Found 224 text files in the folder.
Enter the path to your sample text file: "C:\Users\X1 YOGA\OneDrive\Desktop\Python tasks\text_comparison\sample_text.txt"
How many top similar compositions to analyze? (default 10): 224
Enter the path for the output Excel file (default: results.xlsx): "C:\Users\X1 YOGA\OneDrive\Desktop\Python tasks\text_comparison\Res
ults.xlsx"
Loading sample text from C:\Users\X1 YOGA\OneDrive\Desktop\Python tasks\text_comparison\sample_text.txt...
Sample text loaded successfully (984 characters).
Loading compositions from C:\Users\X1 YOGA\OneDrive\Desktop\Python tasks\text_comparison\students_compositions...
Successfully loaded 224 compositions.
Analyzing compositions... (this may take a moment)
Calculating overall statistics...
Saving results to C:\Users\X1 YOGA\OneDrive\Desktop\Python tasks\text_comparison\Results.xlsx...
Analysis complete! Results saved to C:\Users\X1 YOGA\OneDrive\Desktop\Python tasks\text_comparison\Results.xlsx

The Excel file contains:
- 'Top Similar' - Detailed analysis of the most similar compositions
- 'All Scores' - All compositions ranked by similarity
- 'Summary' - Overall statistics about the analysis

C:\Users\X1 YOGA\OneDrive\Desktop\Python tasks\text_comparison>
```

Figure 6

Python program for correlation analysis



During the code writing for text comparison, some careful considerations were given. They were contrac- tion and capitalization sensitiveness for ensuring accurate semantic aspects, as well as date and number accuracy for including information-bearing items. Regarding the correlation analysis, there were two major statistical tests carried out, they were normal distribution and Spearman correlation (*rho*) tests.

6.4 Procedures

The research procedures were divided into two sub-phases: (1) Text comparison and (2) Correlation analysis. In addition, all of the text and Excel files, as well as the code files needed to belong to the same folder for the Python programs to easily track the available datasets. Regarding the first sub-phase, 224 txt files for students' written compositions and one txt file for a sample text were prepared manually. After that, the newly created Python comparing tool calculated the similarity scores between each students' composition with the sample text and these similarity scores were reported in an Excel file. Subsequently, with the help of an Excel IF function, the most frequently made errors were filtered and reported. Finally, these most frequently made errors were reported to answer the first research question.

In the second sub-phase, two separate Excel files for the previously reported similarity scores and human ratings were manually prepared. Next, these two Excel files were inputted into the second Python program created to check the normal distribution and the correlation of the two datasets. Ultimately, the correlation results were reported to get evidence for the second research question.

7 Results

7.1 Text comparison−The most frequently made errors

As shown in Table 1, the most frequently made errors by students in their legal written compositions belonged to four main categories, which were (1) Past simple and past continuous tense misuse, (2) Reduced relative clause misuse, (3) Misspellings, and (4) Misinformation. These types of errors are coded by the taxonomy proposed by Wu and Garza (2014) consisting of Grammatical, Lexical, Semantic, Mechanics, and Word order (Wu & Garza, 2014).

With two Excel functions of IF and FIN–REPLACE, the occurrences of errors were counted on each student composition, then these occurrences were calculated into percentages. For instance, regarding the first misuse of Past continuous tense, the Excel function of FIN–REPLACE found out there were 23 out of 224 compositions contained this error; next, this occurrence was calculated into percentage and yielded the result of 10.26%.

The most frequently made errors were the misuse of Past simple and Past continuous tenses, with the largest number of compositions containing the misuse of the Past simple form of the verb 'to throw' – 31.69%. Following this number, there were 51 students who wrote different forms of the verb 'to leave' in the sentence 'We didn't leave until the paramedics arrived.', with the percentage of 22.76%. Secondly, 64 out of 224 compositions carried the misuse of Reduced relative clause, with the considerable percentage of 28.57%. Thirdly, regarding misspellings, the percentage of occurrence was approximately 19%; indeed, there were 41 out of 224 compositions which misspelled some important Legal terms. In addition, there were some Information-bearing items throughout 224 compositions. There were 20 students could not include the accurate information of the legal case in their compositions, with the percentage of roughly 9%. Finally, with the help of the IF Excel function, there were 124 out of 224 compositions which contained more than 10 errors, with a noticeable percentage of 55.35%.

Table 1

Types of most frequently made errors

Types of errors	Examples	Occurrences	Percentages
Misuse of tenses (Past simple & continuous)	"... I was jogging..." "... when I noticed..." "... didn't leave until..." "... picked up..." "..... and threw it..."	23 32 51 24 71	10.26% 14.28% 22.76% 10.71% 31.69%
Misuse of Reduced relative clause	"... two young adults shouting..."	64	28.57%
Misspellings	Legal terms	41	18.3%
Misinformation	Date, place, role of writer	20	8.92%
More than 10 errors made		124	55.35%

7.2 Correlation analysis

This study's correlation analysis contained two sub-phases which were the normal distribution test and the correlation test. The entire statistical results are ordinarily grounded on the concept of whether the datasets are normally distributed; as a result, it is a must to carry out this normal distribution test before proceeding with other statistical analysis (Das & Imon, 2016).

According to the aforementioned discussion, the correlation analysis was conducted using the second Python program with the code writing assisted by Claude AI Chatbot. The Shapiro-Wilk results showed that

the Similarity scores dataset was normally distributed, but the Human ratings dataset was not nor- mally distributed, with the results of (W = 0.9832, p = 0.125138), and (W = 0.8463, p = 0.00), respec- tively (see Table 2).

Followed by the normal distribution test of Shapiro-Wilk, the correlation analysis was conducted to check whether the automated comparing tool can reflect human ratings. Since the two datasets were ordinal, where the values were ranked and treated as continuous variables, and to assess the monotonic relationship between the two variables, Spear- man (rho) was chosen (Robitzsch, 2020; Schober et al., 2018). The result revealed a statistically significant correlation between the Similarity scores and Human ratings (*rho* = 0.3042, *p* = 0.000564) (see Table 3 and Figure 7).

Table 2

Shapiro-Wilk results for normal distribution

	Statistic	p-value
Similarity_Normality	0.9832	0.125138
Human_Normality	0.8463	0.000000

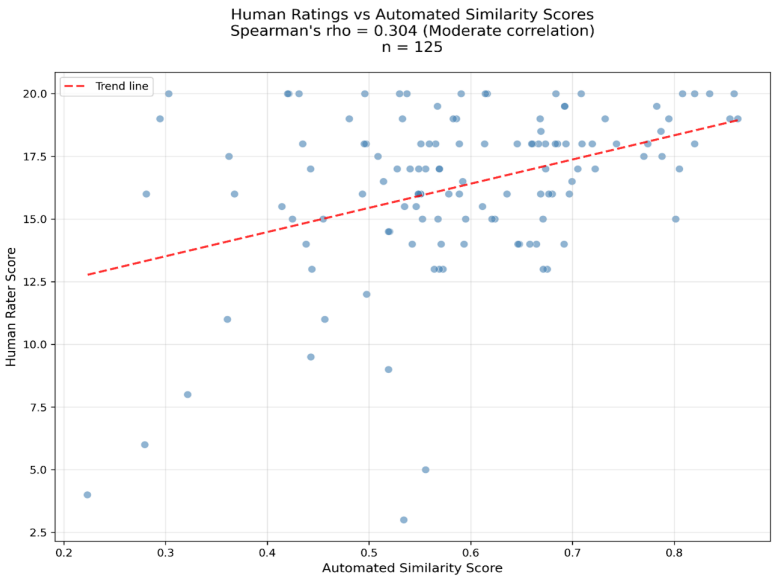
Table 3

Spearman and Pearson correlation results

	Statistic	p-value
Spearman correlation (rho)	0.3042	0.000564
Pearson correlation (r)	0.3886	0.000008

Figure 7

Spearman correlation scatter plot



8 Discussions

8.1 Most frequently made errors

Based on the text comparison results, the central problem arose since the most frequently made errors were reported in the misuse of Past simple and Past continuous tense. The core concept of drafting a witness statement is to narrate the events that happened at the crime scene, and this is served as strong evidence for filing documents in a legal case. These high frequencies of tense misuse raise a question of the Past simple and Past continuous tense teaching and learning in English for Law classrooms. These two hard to learn tenses, to some scientists, need to be taught at the same time to achieve better understanding of their meanings and of how to use them together in a specific context (Anwar et al., 2021; Savova, 2018).

Moreover, there was a relatively high proportion of compositions made the error of Reduced relative clause, with nearly 29%. Students could be confused of this grammar structure in the writing setting where Past simple and Past continuous tenses were commonly used. This might raise another issue of teaching the Relative clause structure together with its shorter versions using present participles or past participles. In this context of writing a witness statement, Law students need to focus on not only how to conjugate verbs, but also their meanings and uses of describing the actions that were happening in the past and another action suddenly happened and interrupted the ongoing action.

Furthermore, there were misinformation errors found in students compositions; indeed, there were inaccurate date and place of the incident, as well as the role of the writer. Besides, as Law students, they are expected to use highly accurate terminology; however, 18.3% of their compositions were full of spelling errors relating to legal terms. Additionally, previous studies proved that using legal terms needs to be paid great attention since these words might carry different orientations and scopes in different legal systems (Teremetskyi et al., 2025). Therefore, focusing on utilizing correct legal terms and identifying the correct legal roles in writing legal documents help Law students prepare for their actual jobs in the future, or these misspelling and misinformation errors are likely to entail negative consequences. Failure to tell what exactly happened, or to give misleading information in a legal case might lead to inconsistencies as this is merely a minor part but crucial in the entire larger process of preparing legal documents. Drafting legal documents requires much more than ‘communication of information and compliance with simple formalities’; as a result, students need to be aware of the accuracy of information and message they use in legal writing (Weinstein & Distler, 1957).

8.2 Correlation between the automated comparing tool and human ratings

The second research question shed light on the practical issue of this research since it tried to validate the newly created comparing tool. If this validation is successful, the tool can be utilized for future ratings. Based on the correlation analysis’ results, there was a moderate correlation between the two datasets. If a composition

received a low similarity score graded by the comparing tool, it had a distinct possibility of getting a low score from the human raters. This is the evidence for claiming that this tool can be used for reducing the teachers' scoring burden. It is also expected to save a lot of time since teachers can initially use this to get reports on the similarity score of each composition compared to the sample text, and teachers can keep track of what the most frequently made errors are. In other words, this tool can be considered as a reference for teachers to have a holistic evaluation followed by a more analytical evaluation. In addition, the results reported in the Excel files might allow teachers to easily filter and find specific errors in students' compositions by using some common functions like FIN–REPLACE or IF.

8.3 Pedagogical implications

According to the study's results, Law students are the most beneficial stakeholders who can have a look back at the errors they made. Students need to be informed of their own errors in their compositions, and this is considered as the act of drawing their initial attention. However, after successfully getting the attention of students on their errors, 'learning of these attended forms might occur based on the premise that attention is what mediates input and intake' (Izumi, 2002). In this stage, teachers-in-charge are the forces playing a more active role in using what had been found in the most frequently made errors for designing more supplementary exercises or remedial lessons; as a result, input is more likely to become intake, and students might prevent themselves in making the same errors later on. These kinds of lessons can be designed in the middle of the semester when students have taken some writing tests in the first half; next, they receive remedies and at the end they can take writing posttests to see whether they are making these errors again, as inspired by many prior studies which have delved into effects of these lessons based on the results of EA (Li, 2019; Zafar, 2016).

Regarding the use of templates in writing tests, although these are effective in helping students write faster, there are some limitations. Legal writing, together with other writing tasks, requires a relatively large number of templates as models to assist students. Nevertheless, this study's results found a substantial proportion of errors made even when students were given clear information and structures were also tactfully suggested. According to some research, templates could make students take stricter control over sentence patterns, vocabulary, and so on; students thus might lose attention to the entire idea or the message they need to convey in their writing. They sometimes merely focus on copying, transforming, or substituting items without being skeptical (Milton & Uddin, 2023). In addition, in the setting of this study, students will attend English for Law 2 module where more difficult writing tasks are awaiting; indeed, they are going to have debate writing or free writing to support a legal proposal. Providing templates needs to be carefully reconsidered when students are encouraged for the more holistic idea of their writing and they are expected to have more room for their free writing in the next module. In terms of using a tailored scoring tool, numerous prior studies have proven that building and validating a specific scoring tool might reduce human rating's

burden, and teachers thus will spend more time giving comprehensive feedback for their students (Ramineni, 2013; Ramineni et al., 2013). Especially, with the discussion of 'error analysis', 'error gravity', and 'diagnostic assessment', teachers can take advantage of these tailored scoring tools not only to document the errors, but also to pay greater attention to errors considered as most serious or impede the idea communication, then teachers are likely to pinpoint 'what aspects of linguistic accuracy' that their students are struggling with (Hughes & Hughes, 2020; Xie, 2019).

9 Concluding remarks

Identifying errors is not only the task of writing teachers but can be a task for students themselves. It is highly recommended that the scope of this study should go beyond reporting on errors made in students compositions to see how students writers 'act upon' these errors. Therefore, students might play an active role as their teachers regarding EA, then researchers can confirm whether students can perceive, store, and apply these information (Sharwood Smith, 1991). Regarding coding procedures, more reliable ratings should be adopted such as having at least two distinct raters and analyze the inter- rater reliability. Additionally, there is another further research idea relating to collecting and analyzing errors that measure 'the writing process', not just working on written products, or the combination of these two processes, the results are expected to be more impressive (Deane, 2013). Next, with the rapid growing of AI technology, creating more scoring, comparing, or analyzing tools tailored for each stage of writing process will be the core development in the near future. Finally, in the setting of ESP teaching and learning at UEH, the idea of EA as well as building and validating scoring tools can be adopted in three other majors with a gigantic number of students, which means larger sample sizes with more reliable results.

References

Anwar, S. R., Chandra, N. E., & Amelia, R. (2021). The ability of students in using simple past tense and past continuous tense at english language education study program batch 2019. *Lingua Educatia Journal*, *3*(4), 309–318.

Ardhy, A. A. S. (2023). *Exploring effective assessment in students' esp writing skill: A conceptual framework. JELTIS*: *Journal of English Language Teaching, Linguistics and Literature Studies*, *3*(2).

Chamberlain, C., Button, A., Dison, L., Granville, S., & Delmont, E. (2004). The role of short answer questions in developing higher order thinking. *Per Linguam*, *20*(2), 28–45.

Chinedu, O., & Ade-Ibijola, A. (2021). Python-bot: A chatbot for teaching python programming. *Engineering Letters*, *29*(1), 25–34.

Das, K. R., & Imon, A. H. M. R. (2016). A brief review of tests for normality. *American Journal of Theoretical and Applied Statistics*, *5*(1), 5–12.

Davar, N. F., Dewan, M. A. A., & Zhang, X. (2025). Ai chatbots in education: Challenges and opportunities. *Information*, *16*, 235.

Deane, P. (2013). On the relation between automated essay scoring and modern views of the writing construct. *Assessing Writing*, *18*, 7–24.

Drid, T. (2018). Ai in language assessment. Language Testing in Asia.

Hughes, A., & Hughes, J. (2020). Testing writing. In *Testing for language teachers* (pp. 87–114). Cambridge University Press.

Hussain, S. S. (2024). Teaching writing to EFL students through error analysis: A cross-sectional study. *Bulletin of the Faculty of Language and Translation*.

Izumi, S. (2002). Output, input enhancement, and the noticing hypothesis [Cited in Han et al., 2008].

Li, C. F. (2019). The application of error analysis in writing based automatic scoring system. *Advances in Social Science, Education and Humanities Research*, *428*, 243–247

Milton, J. D., & Uddin, A. (2023). A case study on "writing difficulties" specifically focusing on "product and process-oriented writing" considering undergraduate students of cox's bazar, Bangladesh. *Scope*, *13*(1), 170–180.

Nunan, D. (2022). *The changing landscape of English language teaching and learning* (Vol. 4). Routledge.

Raimes, A. (1983). *Techniques in teaching writing*. Oxford University Press.

Ramineni, C. (2013). Automated scoring of writing: Validation and reliability. *Educational Measurement: Issues and Practice*, *32*(1), 31–42.

Ramineni, C., et al. (2013). Validating automated scoring. Assessing Writing.

Robitzsch, A. (2020). Why ordinal variables can (almost) always be treated as continuous variables: Clarifying assumptions of robust continuous and ordinal factor analysis estimation methods. Front. Educ. 5:589965. doi: 10.3389/feduc.2020.589965

Savova, L. (2018). Teaching tenses. ELT Journal.

Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation coefficients: Appropriate use and interpretation. *Anesthesia & Analgesia*, *126*(5), 1763–1768.

Sharwood Smith, M. (1991). Speaking to many minds [Cited in Han, 2008].

Tas´ıc, M. (2023). Error analysis in scientific writing: The case of ESP for mechanical engineering. *Innovative Mechanical Engineering*, *2*(1), 1 –12.

Teremetskyi, V., et al. (2025). Legal terminology in esp. *Journal of Legal Studies*.

Ulyani, M. (2024). An error analysis of article system: A case of ESP students. *Yudistira Journal: Publication of Research in Educational Sciences and Language*, *2*(3).

University of Economics, Ho Chi Minh City. (2024). Admission portal. https://tuyensinh.ueh.edu.vn/ Wu, H. P. & Garza, E. V. (2014). Types and attributes of English writing errors in the EFL context−a study of Error

Analysis. *Journal of Language Teaching and Research*, *5*(6), 1256- 1262. doi:10.4304/jltr.5.6.1256-1262

Weinstein, J. B., & Distler, D. H. (1957). Drafting pleading rules. *Columbia Law Review*, *57*(4), 518– 525.

Xie, Q. (2019). Error analysis in legal writing. Journal of Legal Education, 1(1).

Zafar, A. (2016). Error analysis: A tool to improve writing skills of ESL learners. *International Journal of English Linguistics*, *6*(2).

The Impact of Grammarly on Academic Writing Accuracy among Undergraduate EFL Students: A Systematic Review

Hien Thi Phan
National Taiwan Normal University
phanthihien10799@gmail.com

Abstract

Academic writing is an essential skill for undergraduates; however, many students face difficulties with grammatical accuracy. Although modern technology provides various Automated Writing Evaluation (AWE) tools to support this process, students still feel overwhelmed when choosing the most effective tool to use. This systematic review evaluates the effectiveness of Grammarly as an AWE tool in improving academic writing accuracy among EFL undergraduates. Following PRISMA 2020 guidelines, seven studies published between 2016 and 2024 were selected through searches of four databases: ERIC, Scopus, Google Scholar, and ResearchGate. The studies involved undergraduate EFL students and examined Grammarly as the main tool for academic writing. Findings show that Grammarly effectively identifies and corrects surface-level errors, including articles, prepositions, subject–verb agreement, spelling, and punctuation, to improve overall writing accuracy and clarity outcomes. However, Grammarly has limited effectiveness in addressing higher-level writing issues such as organization, coherence, and style. Grammarly's effectiveness is influenced by students' proficiency, the version of the tool used, the extent of feedback uptake, and the complexity of errors. Grammarly is a supplementary tool for improving local grammatical accuracy in EFL academic writing. Therefore, it should not be used as a substitute for teacher feedback, particularly in addressing discourse-level and rhetorical aspects of writing.

Keywords:
Grammarly, EFL Undergraduates, Academic Writing Accuracy, Writing Skills

1 Introduction

With the rapid advancement of technology today, it is necessary to integrate increasingly digital tools for learning English. For EFL learners, writing is an important skill for the development of argument and criticism. Vu (2024) noted that writing is considered one of the most difficult skills because it requires simultaneous control of vocabulary, grammar, sentence structure, and organization. However, L2 learners often focus on lower-level processes such as grammar and vocabulary, which can result in lower-quality writing (Dizon and Gayed, 2021).

At the university level, academic writing is essential. Students are often required to write essays, research reports, and dissertations, which are texts that require clarity, structure, and formal expression. Academic

writing not only demonstrates professional knowledge but also serves as a primary tool for learning and academic communication (Nguyen et al., 2024). However, many students struggle with this skill due to limited understanding of writing conventions, difficulty analyzing prompts, weak sentence writing skills, and challenges in communicating knowledge across contexts (Pineteh, 2014).

To support students, technology, especially Automated Writing Assessment (AWE) tools, has emerged as valuable support tools (Qassemzadeh and Soleimani, 2016; Martínez-Carrasco and Chabert, 2023). AWE systems such as Grammarly, Ginger, and Microsoft Editor are widely used, but choosing the most appropriate and effective tool for academic writing remains a challenge for both students and educators (Bailey and Lee, 2020; Prvinchandar and Ayub, 2014). Among these, Grammarly stands out for its AI-powered feedback and accessibility on grammar, spelling, punctuation, and style. Described as "one of the most accurate and comprehensive language checkers" (Vu, 2024), it has become a popular tool among EFL students (ONeill and Russell, 2019). Nevertheless, research indicates that Grammarly may fail to detect more complex errors, especially those involving semantics or L1 interference, and typically provides less nuanced feedback than a human teacher (Bailey and Almusharraf, 2022).

In L2 writing research, writing quality is commonly evaluated using the Complexity–Accuracy–Fluency (CAF) framework (Skehan, 1998, 2009; Housen et al., 2012). While all three dimensions are important, this review focuses specifically on accuracy, conceptualized as the ability to produce error-free writing that conforms to established grammatical and lexical norms (Ellis & Barkhuizen, 2005). By emphasizing accuracy, the review aims to clarify how Grammarly supports learners in achieving linguistic correctness at the surface level.

Accordingly, this study focuses on investigating the effectiveness of Grammarly in improving writing accuracy, a key component of academic writing quality (Vu, 2024). The study focuses on the tool's ability to detect surface-level grammatical errors (e.g., articles, prepositions, subject-verb agreement) and examines its potential as a supplement to teacher feedback in an EFL context.

2 Literature Review

2.1 Grammarly

Grammarly is an effective writing assistant powered by artificial intelligence (AI) and natural language processing (NLP), and it is also an Automatic Writing Evaluation (AWE) tool (Al-Shaboul et al., 2024; Dizon and Gayed, 2021; Anastasia et al., 2024). Launched in 2009, this tool is considered to be relatively accurate and reliable in detecting writing errors (Fernando and Suryaman, 2022; Liang et al., 2024). Grammarly analyzes multiple aspects of writing, such as syntax, vocabulary, grammar, and clarity, and provides instant and personalized feedback to users (Al-Shaboul et al., 2024).

For EFL students, Grammarly supports autonomy and enhances writing accuracy by providing real-time feedback that helps detect and correct errors (Wang et al., 2024; Geng and Razali, 2022). This feedback reduces repetition errors, promotes self-regulation, and aligns with Schmidt and Strasser (2022) note-taking hypothesis by reinforcing grammar awareness (Anastasia et al., 2024). AWE tools also reduce cognitive load, allowing focus on higher-order writing aspects such as content and organization (Dizon and Gold, 2023).

From a pedagogical perspective, Grammarly benefits teachers by saving time on surface-level corrections. According to Dizon and Gayed (2021), instructors allowed AWE to focus more on providing higher-level feedback, such as content and structure. This shift improves teaching efficiency and feedback quality (Geng and Razali, 2022; Jiang et al., 2020). However, as Wu et al. (2023) pointed out, Grammarly has limitations in identifying deeper discourse issues and needs to be supplemented with teacher feedback.

2.2 This Study

Academic writing in a second language (L2) has shown significant challenges for university students, who are required to address both low-level elements such as grammar and vocabulary, and higher-level aspects such as content and organization (Al-Shaboul et al., 2024; Wu et al., 2023). Many students tend to focus more on surface-level errors than on developing coherence and structure, making grammatical accuracy difficult to achieve (Wu et al., 2023; Vu, 2024). In response to these challenges, AWE tools are increasingly being integrated into instruction to help students improve their writing skills (Liang et al., 2024; Dizon and Gold, 2023).

Despite the wide range of AI-powered tools available today including Grammarly, QuillBot, ChatGPT, and systems like Criterion, students may find it difficult to choose the most appropriate and suitable tool for their academic writing needs (Bailey and Lee, 2020; Calderón and da Cunha Fanego, 2023). Grammarly was chosen as the focus of this study due to its user-friendly interface, widespread use, and positive reception from both learners and instructors (Al-Shaboul et al., 2024; Anastasia et al., 2024). Although Grammarly is known to enhance grammatical accuracy, existing experimental studies are fragmented and lack systematic analysis in university-level writing (Vu, 2024). This study conducts a systematic review of experimental research on Grammarly's impact on EFL students' academic writing accuracy, aiming to provide educators and students with a comprehensive understanding of its effectiveness, the types of errors it addresses, and factors influencing its impact. Grammarly is emphasized as a supplementary, not a replacement, tool in writing instruction (Dizon and Gayed, 2021). This systematic review will address the following research questions:

1. How does Grammarly affect the accuracy of academic writing among undergraduate EFL students?
2. What types of grammatical errors does Grammarly most effectively detect and correct?
3. What factors influence Grammarly's effectiveness in supporting academic writing accuracy in EFL contexts?

3 Methodology

This systematic review was conducted in strict adherence to the PRISMA 2020 (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines to ensure a transparent, replicable, and rigorous study selection process (Page et al., 2021). The review is theoretically grounded in the CAF (Complexity–Accuracy–Fluency) framework, with a specific focus on the Accuracy dimension. Accuracy is conceptualized as the ability to produce linguistically correct writing, encompassing grammar, spelling, punctuation, and sentence structure. By operationalizing accuracy in this manner, the review evaluates Grammarly's effect on linguistic correctness, following established rubrics and frameworks such as those proposed by Weigle (2002) and Ferris (2011).

3.1 Search Strategy and Information Sources

To capture recent developments in Automated Writing Evaluation (AWE) within EFL contexts, a systematic search was performed across four academic databases: ERIC, Scopus, Google Scholar, and ResearchGate. These sources were selected for their comprehensive coverage of peer-reviewed research in educational technology and applied linguistics (Llausas et al., 2024; Nardjes MEFTAH, 2022). The search targeted studies published in English between 2016 and 2024, using the Boolean keyword string: ("Grammarly" OR "AWE") AND ("EFL" OR "L2 learners" OR "university students"). This strategy was designed to retrieve studies examining the intersection of AI-assisted feedback and undergraduate academic writing, ensuring relevance to the research objectives.

3.2 Study Selection

The initial search retrieved 151 studies (ERIC = 41, Google Scholar = 10, ResearchGate = 13, Scopus = 88), which were imported into Microsoft Excel for duplicate removal, leaving 105 studies for screening. A two-stage screening process was then implemented. In the first stage, titles and abstracts were filtered and reviewed against predefined inclusion and exclusion criteria (Table 1), yielding 26 potentially relevant studies. During the second stage, full-text review led to exclusion of 19 studies due to one or more the following reasons: (1) focus on general AWE tools without specific analysis of Grammarly; (2) participant mismatch (non-college or non-EFL learners); (3) Grammarly mentioned but not analyzed separately; or (4) unclear or non-measurable results on writing accuracy.

The screening and coding procedures were conducted by a single reviewer (the author), following a predefined and strictly applied set of inclusion and exclusion criteria to ensure procedural consistency. While single-reviewer screening is common in small-scale systematic reviews, several measures were implemented

to enhance methodological rigor. Specifically, all records were reviewed iteratively at multiple stages, and decisions were cross-checked against the eligibility criteria to minimize selection bias.

To further support consistency, a structured screening and coding protocol was developed and applied throughout the process, creating a transparent audit trail of inclusion and exclusion decisions. NotebookLM (https://notebooklm.google/) was used solely as an organizational tool to assist in data management, categorization, and cross-referencing, without influencing analytical judgments.

Although formal inter-rater reliability was not feasible, reliability was strengthened through repeated coding cycles, systematic cross-checking, and strict adherence to explicit selection criteria. These procedures helped ensure that study selection remained consistent, transparent, and aligned with the research objectives.

3.3 Data Extraction and Management

Data were extracted on study characteristics (authors, year), participants (number, education level), study design (experimental, qualitative, quantitative), duration, and key findings on Grammarly's impact on writing accuracy.

Due to heterogeneity in designs, participants, and outcome measures, a meta-analysis was not feasible. Consequently, a narrative synthesis was employed, allowing for a conceptual interpretation of the evidence. This approach supports a nuanced discussion of how Grammarly facilitates improvements in writing accuracy across diverse EFL contexts, emphasizing thematic patterns such as error treatability, learner proficiency, and feedback engagement without overgeneralizing results from a limited evidence base.

To ensure methodological rigor, study quality was evaluated using four specific indicators: (1) adequacy of research design, (2) clarity and validity of accuracy measurement, (3) completeness of participant description, and (4) transparency of data analysis procedures. Each study was systematically assessed against these indicators using a structured evaluation framework.

Table 1

Inclusion and Exclusion Criteria

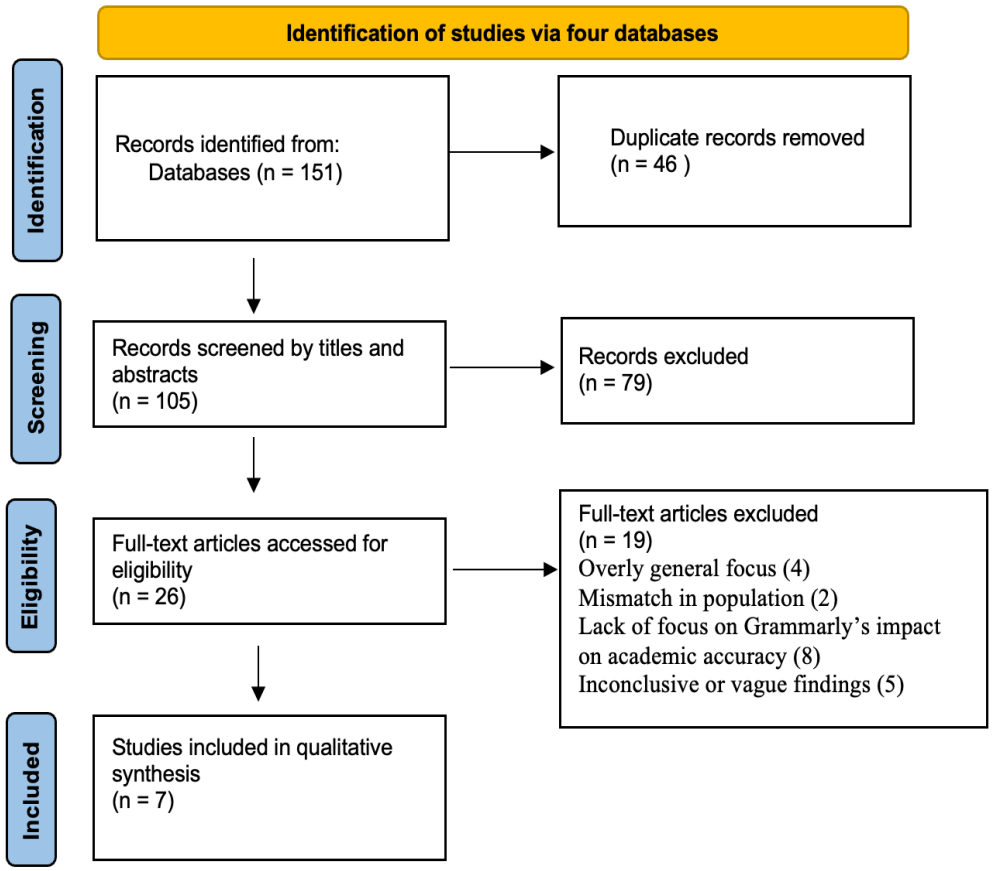
Inclusion criteria	Exclusion criteria
- Studies involve EFL undergraduates.	- Studies involving non-undergraduate populations
- Studies use Grammarly tool and examines its impact on academic writing accuracy.	- Studies without specific focus on writing accuracy
- Studies published between 2016 and 2024.	- Studies published outside 2016-2024 range
- The study is written in English.	- The study is not written in English
- The full text of the study is available for review.	- The study with no full-text access

Based on this assessment, study quality was operationalized as high, moderate, or low, with only studies meeting a minimum threshold retained for synthesis. This operationalization ensured a consistent, transparent, and methodologically sound evaluation across all studies.

The PRISMA process flow diagram was developed as a standard method to demonstrate how to search, evaluate, and integrate studies in systematic reviews and meta-analyses. This tool increases transparency and understanding by describing each step in detail, starting from the initial search stage to the final selection of eligible studies. The selection process is carried out through four screening steps (Llausas et al., 2024). First, 151 studies were identified through searches across four databases. After removing 46 duplicate studies, 105 studies remained for the screening stage, where inclusion and exclusion criteria were applied. In this stage, 79 records were excluded as irrelevant, resulting in 26 studies. After reading the full text of these 26 studies, 19 studies were further excluded due to not meeting the specific research requirements, ultimately leaving 7 studies that were selected most suitable and included in this systematic literature review. The process is demonstrated in Figure 1.

Figure 1

PRISMA flowchart



4 Results and Discussion

The findings from the reviewed studies suggest that Grammarly's effectiveness in improving academic writing accuracy for EFL undergraduates depends on the nature of errors, learner proficiency, and engagement with feedback. Table 2 summarizes the key features of the studies included in this review, including research design, participant information, data collection methods, and study duration. The discussion is organized into three thematic areas: Error Treatability and Writing Accuracy, Learner Proficiency and Interaction with Grammarly, and Feedback Engagement and Revision Processes.

Table 2

Features of studies

Author (Year)	Research Method	Data Collection Method	Research Period	No. of Undergraduates
Almusharraf & Alotaibi (2023)	Quantitative	Midterm essays, error analysis (human & Grammarly)	One semester	197
Bailey & Lee (2020)	Quantitative	Corpus analysis	16 weeks	163
Miranty et al. (2023)	Experimental (Quantitative)	Questionnaires, writing tests	14 weeks	35
Dizon & Gayed (2021)	Quantitative	Composition tasks, analysis of writing reports	8 weeks	31
Dewi (2022)	Qualitative	Questionnaires, interviews	Unspecified duration	75
Abu Guba et al. (2024)	Quantitative	Essays, Grammarly analysis	One semester	60
Thi N.K. et al. (2023)	Qualitative	Questionnaires, feedback (teacher & Grammarly), revised texts	14 weeks	31

4.1 Error Treatability and Writing Accuracy

The review evidence suggests that Grammarly's impact on writing accuracy is closely linked to the treatability of errors (Error Treatability). According to the CAF framework (Complexity–Accuracy–Fluency), Grammarly appears to primarily influence surface-level accuracy rather than higher-order complexity or fluency in L2 writing.

Table 3 summarizes the types of errors that Grammarly was found to detect most effectively across the included studies.

Table 3

Types of errors Grammarly detects and corrects most effectively in the included studies

CE	No.	SGE	No.	SE	No.	SWUE	No.
Spelling	7	Verb Errors	5	Fragment Sentences	4	Word Choice Errors	7
Punctuation	6	Article/Determiner Errors	5	Run-on Sentences	4	Wordiness/Redundancy	5
Capitalization	5	Noun Form/ Number Errors	5	Comma Splices	4	Informal/Slang Usage	4
Grammatical	7	Pronoun Errors	3			Tone Issues	3
Sentence structure	5	Preposition Errors	5			Vocabulary Repetition (Lack of Variety)	4
		Conjunction Errors	3				
		Modifier Errors	4				

Notes. CE = Correctness Errors, SGE = Specific Grammatical Errors, SE = Sentence-level Errors, SWUE = Style/Word Usage Errors, No. = Number of studies

Across studies, Grammarly appears most effective in addressing treatable errors that follow clear grammatical rules, such as articles, verb forms, subject-verb agreement, prepositions, spelling, and punctuation (Dizon & Gayed, 2021; Abu Quba et al., 2024; Bailey & Lee, 2020). This suggests that AI-assisted writing tools can support learners in attending to systematic linguistic patterns and may also enhance lexical diversity and word choice through predictive suggestions (Dizon & Gayed, 2021); Almusharraf & Alotaibi, 2023). Students' perceptions further indicate that Grammarly helps identify and correct grammar, punctuation, spelling, and word-choice errors, which aligns with theories of feedback uptake emphasizing that immediate, form-focused feedback scaffolds attention to linguistic accuracy (Dewi, 2022).

However, Grammarly seems less effective for higher-order or untreatable errors, including complex sentence structures, cohesion, style, and subtle semantic choices (Almusharraf & Alotaibi, 2023; Thi et al., 2023; Bailey & Lee, 2020). Errors such as run-on sentences, comma splices, logical inconsistencies, or discourse-level issues were often missed or incorrectly flagged. These finding aligns with L2 writing development theory, which posits that learners typically acquire control over basic grammatical structures before mastering discourse-level features. Thus, while Grammarly is valuable for mechanical and local error correction, teacher guidance remains essential for coherence, rhetorical quality, and higher-order writing skills.

4.2 Learner Proficiency and Interaction with Grammarly

Learner proficiency appears to moderate the effectiveness of Grammarly. Evidence suggests that low-proficiency students benefit substantially, showing improvements in treatable errors and basic writing mechanics (Abu Quba et al., 2024). For these learners, automatic feedback scaffolds their attention to grammar and reduces surface-level inaccuracies, thereby enhancing post-intervention performance.

The version of Grammarly used may also influence these outcomes. Research indicates that the premium version, which checks over 400 error types, provides more comprehensive feedback than the free version, which has more limited errors detection capabilities (Almusharraf & Alotaibi, 2023). Learners using the free version may receive fewer suggestions, potentially limiting improvements in grammatical accuracy, lexical diversity, and word-choice refinement. Furthermore, the cost of the premium version may restrict access for some students, which can partially explain variability in observed effectiveness across studies.

In contrast, higher-proficiency learners engage more selectively with automated feedback. They may question the accuracy or relevance of suggestions, leading to partial uptake or rejection of corrections, especially for high-order writing tasks (Thi et al., 2023). This pattern may be explained by varying levels of metalinguistic awareness: less proficient learners rely on external feedback, whereas advanced learners critically evaluate and integrate suggestions selectively. Consequently, Grammarly's effectiveness is not uniform but appears to depend on learners' ability to interpret and apply feedback appropriately.

This proficiency-dependent interaction underscores the importance of contextualizing AWE use, ensuring that learners at different skill levels can engage meaningfully with automated feedback.

4.3 Feedback Engagement and Revision Processes

The effectiveness of Grammarly also depends on how learners engage with feedback. Studies indicate that receiving automated feedback does not automatically translate into correct revisions. Detailed analyses show high rates of "No Revision" (NR) and "Incorrect Revision" (IR), alongside lower rates of "Correct Revision" (CR). For instance, in Thi et al. (2023), punctuation errors showed CR=27.2% and NR=70.9%, subject-verb agreement CR=20% and NR=80%, and pronouns CR=15.4% and NR=84.6%.

These patterns suggest that learners may misinterpret, distrust, or fail to apply feedback correctly, particularly in complex or ambiguous contexts. Students appear more receptive to form-focused feedback (grammar, spelling, punctuation) than meaning-focused feedback, consistent with Grammarly's design, but uptake remains inconsistent (Thi et al., 2023; Dewi, 2022).

Learners' trust in automated feedback also plays a role. Without awareness of tool limitations – including false positives or inappropriate suggestions – students may ignore valid feedback or apply it incorrectly. Therefore, Grammarly's effectiveness depends not only on its error detection capability but also on learners'

cognitive engagement, metalinguistic knowledge, and reflective revision practices. Instructor scaffolding remains crucial to help learners translate automated suggestions into genuine improvements in writing accuracy.

Synthesizing the reviewed studies suggests that Grammarly is most effective for treatable, surface-level errors, particularly among lower-proficiency learners. Its impact diminishes for higher-order errors involving cohesion, style, and discourse-level organization, and is moderated by learner proficiency and feedback engagement. While Grammarly supports mechanical accuracy and lexical diversity, it cannot replace teacher guidance in fostering coherent, rhetorically effective L2 writing.

5 Conclusion

Overall, research shows widespread agreement that Grammarly is effective in improving the accuracy of academic writing for EFL undergraduates, primarily by detecting and correcting surface errors. The tool does a particularly good job of identifying common issues such as article usage, subject-verb agreement, prepositions, punctuation, and spelling errors (Bailey and Lee, 2020; Dizon and Gayed, 2021). For lower-level students, Grammarly helps reduce error frequency and build grammatical awareness through real-time feedback. However, the actual impact of Grammarly depends on a number of factors. The level of the students influences how effectively they interpret and apply feedback. Lower-level students may benefit from basic corrections, while higher-level students may find the suggestions insufficient or too simplistic. Effectiveness also varies depending on the version used (free vs. premium), which determines the range of errors detected.

Importantly, Grammarly's ability to improve accuracy is limited by the students' ability to understand and perceive the feedback provided. Research shows that some students ignore or misunderstand the suggestions, reducing the tool's pedagogical value. Furthermore, Grammarly has not been shown to be effective in addressing complex grammatical structures, sentence-level coherence, paragraph organization, and rhetorical appropriateness. The accuracy and appropriateness of the suggestions also vary by error type.

For these reasons, Grammarly should be used as a supplement, not a replacement, for teacher feedback. Teachers still play an essential role in addressing higher-level writing concerns and engaging students in critical reflection with automated feedback. Being aware of Grammarly's strengths and limitations is key to using the tool effectively in academic writing instruction.

5.1 Pedagogical Implications

The findings of this review have several implications for writing instruction in EFL contexts. First, Grammarly can serve as a useful supplementary tool for both students and instructors. For students, it offers immediate feedback on surface-level errors, enabling more independent and confident revision. For instructors, Grammarly can reduce the time spent on correcting errors, allowing them to focus on higher-order aspects such as argument development, coherence, and rhetorical structure.

Additionally, the use of Grammarly in writing classrooms could foster student autonomy and increase grammar awareness. Teachers can incorporate Grammarly into writing workshops or revision stages, using it for classroom discussions on common error types. However, to ensure effective integration, students should be trained on how to interpret Grammarly's feedback and recognize its limitations.

5.2 Limitations and Future Research

Despite the systematic approach employed in this review, several limitations should be acknowledged. First, the evidence base is relatively small, comprising only seven studies, which limits the generalizability of findings. While these studies provide valuable insights into EFL undergraduate contexts, the results should be interpreted as indicative of general trends rather than definitive evidence of Grammarly's long-term effectiveness.

Second, the screening and coding were conducted by a single researcher. Although the structured application of PRISMA 2020 guidelines and the use of NotebookLM helped reduce subjectivity, the absence of formal inter-rater validation may affect the reliability of study selection.

Third, substantial heterogeneity in study designs and assessment tools precluded a meta-analysis, necessitating a narrative synthesis. While this approach allows for conceptually rich interpretation, it limits the ability to quantify effect sizes for error reduction. Finally, the reviewed studies predominantly examined short-term interventions. Future research should adopt longitudinal designs to determine whether immediate gains in grammatical accuracy translate into sustained L2 writing development and to explore how learner proficiency influences engagement with automated feedback.

References

The reviewed articles are labelled with an asterisk (*)

*Abu Guba, M. N., Awad, A., & Abu Qub'a, A. (2024). Grammarly in teaching writing to EFL learners at low levels: How useful is it? *World Journal of English Language, 14*(3), 1–9. https://doi.org/10.5430/wjel.v14n3pl

Al-Shaboul, I. A., Al Rousan, R. M., Kalsoom, T., & A-Awawdeh, N. (2024). A critical examination of how AI-driven writing tools have impacted the content, style, and organization of foreign language undergraduates' writing: A survey of lecturers. *Research Journal in Advanced Humanities, 5*(4), 31–48. https://doi.org/10.58256/pjf6en83

*Almusharraf, N., & Alotaibi, H. (2023). An error-analysis study from an EFL writing context: Human and automated essay scoring approaches. *Technology, Knowledge and Learning, 28*(3), 1015-1031. https://doi.org/10.1007/s10758-023-09640-w

Anastasia, W., Murtisari, E., & Isharyanti, N. (2024). Lower-proficiency EFL students' use of Grammarly in writing: Behavior, cognition, and affect. *The JALT CALL Journal, 20*(1)*,* 1-25.

*Bailey, D., & Lee, A. R. (2020). An exploratory study of Grammarly in the language learning context: An analysis of test-based, textbook-based and Facebook corpora. *TESOL International Journal, 15*(2)*,* 4-27.

Bailey, D. R., & Almusharraf, N. (2022). A structural equation model of second language writing strategies and their influence on anxiety, proficiency, and perceived benefits with online writing. *Education Information Technologies, 27*(8)*,* 10497-10516.

Calderón, S. S., & Da Cunha Fanego, I. (2023). Developing autonomy in English writing skills: a study on EFL learners' selection criteria of open access and online self-correction tools. *CÍRCULO de Linguística Aplicada a la Comunicación, 95*(1)*,* 95-105.

*Dewi, U. (2022). Grammarly as automated writing evaluation: Its effectiveness from EFL students' perceptions. *Lingua Cultura, 16*(1)*,* 155-161.

*Dizon, G., & Gayed, J. M. (2021). Examining the impact of Grammarly on the quality of mobile L2 writing. *The JALT Call Journal, 17*(2)*,* 74-92. https://doi.org/10.29140/jalt-call.v17n2.336

Dizon, G., & Gold, J. (2023). Exploring the effects of Grammarly on EFL students' foreign language anxiety and learner autonomy. *JALT CALL Journal, 19*(3)*,* 299-316.

Fernando, E. K., & Suryaman, M. (2022). Appropriating grammar as a tool in writing activities: student's perception. *Journal of English Language Education, 7*(1)*,* 1-8.

Ferris, D. (2011). *Treatment of error in second language student writing*. University of Michigan Press.

Geng, J., & Razali, A. B. (2022). Effectiveness of the Automated Writing Evaluation program on improving undergraduates' writing performance. *English Language Teaching, 15*(4)*,* 49-60.

Jiang, L., Yu, S., & Wang, C. (2020). Second language writing instructors' feedback practice in response to automated writing evaluation: A sociocultural perspective. *System, 93,* 102302.

Liang, J., Huang, F., & Teo, T. (2024). Understanding Chinese university EFL learners' perceptions of AI in English writing. *International Journal of Computer-Assisted Language Learning Teaching, 14*(1)*,* 1-16.

Llausas, S. M., Ruiz, E., Ayucan, S. M., & Evardo Jr, O. J. (2024). A systematic literature review on the use of Grammarly in improving the writing skills of ESL/EFL students. *International Journal of Multidisciplinary: Applied Business and Education Research, 5*(9), 3507–3516. https://doi.org/10.11594/ijmaber.05.09.10

Martínez-Carrasco, R., & Chabert, A. (2023). Writing on steroids? Accuracy of automatic corrective feedback in L2 competence development. *Bellaterra Journal of Teaching Learning Language Literature, 16*(1), 1142.

*Miranty, D., Widiati, U., Cahyono, B. Y., & Sharif, T. (2023). Automated writing evaluation tools for Indonesian undergraduate English as a foreign language students' writing. *International Journal of Evaluation Research in Education, 12*(3)*,* 1705-1715.

Nguyen, H. H. B., Ngoc, H. H. B., & Dan, T. C. (2024). EFL students'perceptions and practices of using Chatgpt for developing English argumentative essay writing skills. *European Journal of Alternative Education Studies, 9*(1), 168–216. https://doi.org/10.46827/ejae.v9i1.5341.

O'Neill, R., & Russell, A. (2019). Stop! Grammar time: University students' perceptions of the automated feedback program Grammarly. *Australasian Journal of Educational Technology, 35*(1), 42–56. https://doi.org/10.14742/ajet.3795.

Page, M. J., Mckenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., & Brennan, S. E. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372,* n71.

Pineteh, E. A. (2014). The academic writing challenges of undergraduate students: A South African case study. *International Journal of Higher Education, 3*(1)*,* 12-22.

Prvinchandar, S., & Ayub, A. F. M. (2014). Comparison of the effectiveness of Stylewriter and microsoft word computer software to improve English writing skills. *English Language Teaching, 7*(1)*,* 93-102.

Qassemzadeh, A., & Soleimani, H. (2016). The impact of feedback provision by Grammarly software and teachers on learning passive structures by Iranian EFL learners. *Theory and Practice in Language Studies, 6*(9)*,* 1884 –1894. https://doi.org/10.17507/tpls.0609.23.

Schmidt, T., & Strasser, T. (2022). Artificial intelligence in foreign language learning and teaching: A CALL for intelligent practice. *Anglistik: International Journal of English Studies, 33*(1)*,* 165-184.

*Thi, N. K., Nikolov, M., & Simon, K. (2023). Higher-proficiency students' engagement with and uptake of teacher and Grammarly feedback in an EFL writing course. *Innovation in Language Learning and Teaching, 17*(4)*,* 690-705.

Vu, P. T. (2024). Using Grammarly in enhancing students' grammatical accuracy in English writing: A case at a private college in Hanoi. In *Proceedings of the AsiaCALL International Conference* (Vol. 4, pp. 81–96)

Wang, S., Xu, T., Li, H., Zhang, C., Liang, J., Tang, J., Yu, P. S., & Wen, Q. (2024). Large language models for education: A survey and outlook. *arXiv preprint arXiv:18105*.

Weigle, S. C. (2002). *Assessing writing*. Cambridge University Press.

Wu, H., Wang, W., Wan, Y., Jiao, W., & Lyu, M. (2023). ChatGPT or Grammarly? Evaluating ChatGPT on grammatical error correction benchmark. *arXiv preprint arXiv:13648*.

Investigating the Relationship Between Students' Attitudes Towards English Pronunciation and Performance

Kanae Ohashi

Ritsumeikan University

Kohashi5@fc.ritsumei.ac.jp

Abstract

This study examined the relationships between Japanese EFL learners' attitudes toward pronunciation, pronunciation awareness, and pronunciation performance as measured by an automatic speech recognition (ASR) system. A total of 202 university students completed a questionnaire on pronunciation-related attitudes, awareness, and learning behaviors. From the collected speech data, 40 read-aloud recordings (10 from each proficiency level) were selected for ASR-based analysis using the ELSA Speech Analyzer. Segmental and suprasegmental pronunciation scores were analyzed separately, and Pearson's correlation analyses were conducted. The results revealed distinct patterns for segmental and suprasegmental pronunciation performance. Segmental scores were strongly associated with learners' active motivation, confidence, awareness of specific segmental features, and engagement in singing English songs. In contrast, suprasegmental scores were most strongly related to learners' awareness of core prosodic features, such as intonation and word grouping, as well as enjoyment of listening to English songs. Passive beliefs about the importance of pronunciation showed limited associations with performance. These findings suggest that pronunciation development is closely linked to active engagement and domain-specific awareness rather than belief alone. Despite limitations related to sample size and the use of ASR-based assessment, this study highlights the pedagogical potential of ASR for examining pronunciation development in EFL contexts.

Keywords

pronunciation awareness, segmental and suprasegmental features, ASR (Automatic Speech Recognition)

1 Introduction

1.1 Background of the Study

In an era of globalization characterized by an unprecedented pace of progress, and with English functioning as an international language, intelligible pronunciation has become increasingly important for effective oral communication. Researchers and educators have consistently identified mispronunciation as a major source of communication breakdown (e.g., Jenkins, 2000).

Despite this recognition, pronunciation instruction in Japan has traditionally received less emphasis than other components of language learning, such as grammar and lexis. As a result, many Japanese learners of English as a foreign language (EFL) experience difficulties in being understood in international communicative contexts.

There have been several studies investigating Japanese EFL learners' attitudes toward pronunciation and their learning experiences. These studies have shown that, although many learners recognize the importance of pronunciation, a substantial number have not received sufficient instruction in this area (e.g., Armand & Tanaka, 2001; Ohashi, 2018).

1.2 The Phonological Challenges Faced by Japanese EFL Learners

As a result of first language (Japanese) transfer, Japanese EFL learners face a range of phonological challenges. At the suprasegmental level, these challenges include (1) vowel insertion, which leads to an increased number of syllables, and (2) weak, flat stress patterns at both the word and sentence levels.

At the segmental level, Japanese learners often have difficulty distinguishing between certain consonants, such as /s/ vs. /ʃ/, /t/ vs. /tʃ/, /b/ vs. /v/, and /l/ vs. /r/, as well as producing word-initial /w/ and /j/. With regard to vowels, Ehrlich and Avery (2013) note that contrasts between tense and lax vowels (*e.g.*, /i/ vs. /ɪ/) and among low and mid vowels (*e.g.*, /ɛ/, /æ/, /ɑ/, and /ʌ/) are particularly challenging for Japanese learners.

1.3 Purpose of the Study

Given the current situation of Japanese EFL learners, this study has three main purposes. First, it aims to investigate Japanese university students' perceptions of their own English pronunciation. Second, it seeks to compare these self-perceptions with their actual pronunciation performance. Specifically, the study examines whether learners' perceived difficulties with particular pronunciation features are reflected in their measured performance. Finally, this study aims to discuss the role of pronunciation intelligibility in contemporary English education in Japan.

2 Method

2.1 Participants

The participants of this study were 202 non-English major freshmen, who took the author's and two other instructors' English CALL (Computer Assisted Language Learning) classes. In the class, students acquired research skills, information and communication technology (ICT) literacy, and pronunciation skills through computer-assisted instruction. The students' proficiency levels were determined based on the university's English placement test used for course placement, which roughly corresponds to CEFR levels ranging from A1 to B2.

2.2 Procedure

The students were first asked to complete a questionnaire consisting of 18 items designed to elicit their general attitudes and awareness toward English pronunciation. The questionnaire included items addressing learners' attitudes toward English pronunciation, their awareness of specific pronunciation features *(e.g.*, whether they pay attention to producing the /l/ sound accurately), and their daily activities related to English pronunciation, such as listening to or singing English songs (see Appendix A). The questionnaire was administered in October 2023.

Prior to completing the questionnaire, the students were informed that their responses would not affect their course grades. Only the data from students who provided informed consent for their responses to be used for research purposes were included in the analysis.

Along with the questionnaire, the students were asked to submit a recorded reading-aloud audio file based on the text provided in Appendix B for analysis using an automatic speech recognition (ASR) system, ELSA Speech Analyzer. Of all the recordings collected, 10 samples from each proficiency level were selected for ASR analysis (N = 40). The ASR system evaluated students' pronunciation performance in terms of segmental and suprasegmental features, which were reported as percentage scores. Descriptive statistical analyses were then conducted to identify general tendencies in the learners' actual pronunciation performance. Finally, Pearson's correlation analyses were conducted to examine the relationships between students' pronunciation-related attitudes and behaviors and their ASR-based pronunciation scores. All statistical analyses were conducted using IBM SPSS Statistics for Windows (Version 29).

2.3 Automatic Speech Recognition (ASR)

As noted above, pronunciation analysis was conducted using the ELSA Speech Analyzer, an automatic speech recognition (ASR)-based pronunciation assessment tool developed by ELSA Corp. (available at https://speechanalyzer.elsaspeak.com/).The software is designed to evaluate spoken English across five dimensions: (1) pronunciation, (2) intonation, (3) fluency, (4) grammar, and (5) vocabulary.

According to the ELSA development team (Anguera et al., 2023), the pronunciation score reflects the accuracy of phonetic production relative to native-speaker models. Drawing on ASR technology, the system detects phonetic deviations in learners' speech that may affect intelligibility and generates corresponding feedback. In this component, an automatic transcription of the speech is also provided. In the present study, the pronunciation score was treated as a segmental measure.

The intonation score is derived from analyses of pitch, energy, and word prominence. Pitch and energy are examined in terms of their appropriateness and variability, while word prominence is evaluated by comparing expected stress patterns with actual acoustic realizations. Both overemphasized and underemphasized words contribute to the prominence score. In the present study, intonation scores were treated as suprasegmental measures.

Because the speech samples analyzed in this study consisted of read-aloud speech rather than spontaneous

production, the remaining dimensions, that is, fluency, grammar, and vocabulary, were not included in the analysis.

3 Results

3.1 Descriptive Statistics

Table 1 - 3 present the results obtained by the questionnaire (See Appendix A). Table 1 presents the students' attitudes toward English and English pronunciation. As shown in the table, students at higher proficiency levels generally reported more positive attitudes toward English. However, learners across all proficiency levels recognized the importance of pronunciation and expressed a desire to improve their pronunciation, regardless of their overall proficiency.

Table 1

Attitudes toward English

Level	Likes Learning English (*M, SD*)	Importance of Pronunciation (*M, SD*)	Desire to Improve Pronunciation (*M, SD*)
1	4.18 (1.08)	5.36 (0.75)	5.45 (0.58)
2	3.73 (0.98)	4.95 (0.91)	5.26 (0.77)
3	3.32 (1.02)	5.03 (0.80)	4.91 (0.91)
4	3.00 (0.88)	5.14 (0.77)	4.93 (0.84)
Total	3.66 (1.04)	5.02 (0.84)	5.20 (0.78)

Note. Levels are ranked from 1 (highest) to 4. Scores range from 1 (weakest agreement) to 6 (strongest agreement).

N = 202

Table 2

Pronunciation Awareness for Suprasegmentals

Level	Linking (*M, SD*)	Intonation *(M, SD)*	Chunking by Meaning (*M, SD*)	Content vs Function words (*M, SD*)	Word Stress (*M, SD*)
1	3.95 (0.94)	3.95(0.91)	3.55 (0.97)	3.27 (0.90)	3.45 (0.99)
2	3.64 (0.89)	3.58(0.97)	3.76 (0.89)	3.34 (0.96)	3.40 (0.92)
3	3.12 (0.92)	2.94 (0.88)	3.06 (0.92)	2.65 (0.84)	2.94 (0.94)
4	2.86(0.86)	3.00 (0.83)	3.00 (0.85)	2.50 (0.76)	2.86 (0.89)
Total	3.53(0.95)	3.48 (0.93)	3.56 (0.91)	3.16 (0.92)	3.29 (0.94)

Note. Levels are ranked from 1 (highest) to 4. Scores range from 1 (weakest agreement) to 6 (strongest agreement).

N = 202

Table 2 presents the extent to which students were aware of suprasegmental-related features. Overall, students at higher proficiency levels demonstrated greater awareness across most items than those at lower proficiency levels. It is also notable that awareness of the distinction between content and function words was lower than that of other pronunciation features.

Table 3

Pronunciation Awareness for Segmentals

Level	Pronunciation of /r/ (*M*, *SD*)	Pronunciation of /l/ (*M*, *SD*)	Pronunciation of /θ/ (*M*, *SD*)	Non-Japanese Consonants (*M*, *SD*)	Non-Japanese Vowels (*M*, *SD*)
1	3.18 (0.87)	3.64 (0.90)	3.91 (0.88)	3.27 (0.88)	3.14 (0.89)
2	3.20 (0.92)	3.15 (0.84)	3.45 (0.92)	3.07 (0.83)	3.00 (0.86)
3	3.03 (0.90)	3.18 (0.88)	3.27 (0.91)	2.88 (0.85)	2.97 (0.84)
4	2.64 (0.79)	2.57 (0.75)	3.00 (0.80)	2.71 (0.74)	2.64 (0.77)
Total	3.13 (0.88)	3.17 (0.87)	3.44 (0.88)	3.04 (0.85)	2.99 (0.85)

Note. Levels are ranked from 1 (highest) to 4. Scores range from 1 (weakest agreement) to 6 (strongest agreement).

N = 202

Table 3 indicates students' awareness of segmental features. The questionnaire items examined whether learners paid attention to these features during speech production. For example, students were asked whether they attended to accurate production of the /r/ sound when speaking. The results show that vowel-related features were the least attended to, regardless of proficiency level.

Table 4 presents students' pronunciation performance scores obtained through automatic speech recognition (ASR) analysis. Overall, pronunciation performance tended to align with proficiency levels. Notably, segmental scores were generally lower than suprasegmental scores across all proficiency levels.

Based on the descriptive statistics, three overall trends were identified. First, students across all proficiency levels demonstrated generally positive attitudes toward pronunciation. Second, a clear relationship was observed between proficiency level and pronunciation awareness. Third, students showed greater awareness of basic suprasegmental features than of more complex segmental features.

Table 4

ASR-Based Language Proficiency Report by Level

Level	Metric	Overall	Segmentals	Suprasegmentals
1	*M*	80.50	65.70	81.80
	n	10	10	10
	SD	6.13	16.57	8.16
2	*M*	77.60	57.20	79.60
	n	10	10	10
	SD	3.57	11.06	9.20
3	*M*	77.18	63.91	74.73
	n	11	11	11
	SD	3.89	7.94	8.83
4	*M*	71.22	54.22	66.33
	n	9	9	9
	SD	5.86	12.67	11.58
Total	*M*	76.77	60.50	75.83
	n	40	40	40
	SD	5.79	12.76	10.79

Note. Levels are ranked from 1 (highest) to 4. In ASR, the score is shown as a percentage. *N*=40

3.2 Correlations Between ASR-Based Pronunciation Scores and Related Attitudes

Because multiple correlations were examined, there is a possibility of increased Type I error (false positives). Therefore, the results should be interpreted with caution. Tables 5 and 6 present the correlations between segmental and suprasegmental scores obtained through ASR analysis, respectively. Both sets of scores were normally distributed. Visual inspection of scatterplots suggested that the relationships between variables were approximately linear. Therefore, Pearson's correlation analyses were conducted.

Table 5 presents the correlations between students' segmental pronunciation performance scores and related attitudinal variables. Specifically, the analysis examined whether students who reported paying attention to particular segmental features (*e.g.*, the /r/ sound) demonstrated higher segmental performance scores. The results revealed significant positive correlations between segmental performance and learners' desire to improve their pronunciation, awareness of the consonants /r/ and /l/, self-reported confidence in producing consonants and vowels accurately, and engagement in singing English songs.

Table 5

Correlations between Pronunciation (Segmentals) Scores and Related Attitudes

Variable	*r*	*p*	Effect Size (R^2)
Likes Learning English	.43	.006	.182 (Medium)
Importance of Pronunciation	.25	.125	.061 (Small)
Desire to Improve Pronunciation	.66**	<.001	.432 (Large)
/ r /	.60**	<.001	.362 (Large)
/ l /	.57**	<.001	.321 (Large)
/ θ / (th sound)	.40	.010	.161 (Medium)
Consonants	.56**	<.001	.316 (Large)
Vowels	.57**	<.001	.321 (Large)
Confidence	.40	.010	.161 (Medium)
Listens to English Songs	.34	.033	.114 (Medium)
Sings English Songs	.60**	<.001	.356 (Large)
Likes English songs	.43	.006	.182 (Medium)

Note. *N*=40

Table 6 presents the correlations between students' suprasegmental pronunciation performance scores and related attitudinal variables. Significant positive correlations were found with students' self-reported confidence in intonation and word grouping, as well as with their liking of English songs.

Table 6

Correlations between Intonation Scores (Suprasegmentals) and Attitudes

Variable	*r*	*p*	Effect Size (R^2)
Likes Learning English	.31	.056	.093 (Small)
Importance of Pronunciation	.27	.095	.072 (Small)
Desire to Improve Pronunciation	.29	.073	.081 (Small)
Linking	.45*	.003	.200 (Medium)
Intonation	.71**	<.001	.504 (Large)
Word group	.57**	<.001	.319 (Large)
Content words	.27	0.09	.075 (Small)
Word stress	.25	.121	.061 (Small)
Confidence	.25	.121	.062 (Small)
Listens to English Songs	.27	.093	.072 (Small)
Sings English Songs	.06	.697	.004 (Small)
Likes English songs	.40**	.011	.158 (Medium)

Note. *N*=40

Taken together, the results from Tables 5 and 6 show clear differences in how segmental and suprasegmental pronunciation performance relates to learners' attitudes and awareness. Segmental pronunciation scores were positively associated with learners' desire to improve pronunciation, awareness of specific consonants and vowels, self-reported confidence, and active engagement with English songs, particularly singing. In contrast, general beliefs about the importance of pronunciation showed weak or non-significant relationships with segmental performance.

Suprasegmental pronunciation scores, on the other hand, were most strongly associated with learners' awareness of core prosodic features, including intonation, word grouping, and linking. While most attitudinal variables showed weak or non-significant relationships with suprasegmental performance, learners' enjoyment of English songs showed a significant positive association.

4 Discussions and Implications

Based on the results, four main implications for effective pronunciation instruction can be identified. First, learners' awareness of pronunciation features was found to align with their actual pronunciation performance. In other words, students who reported greater awareness of both segmental and suprasegmental features tended to demonstrate higher levels of pronunciation performance. This finding suggests that raising learners' awareness of pronunciation features plays a crucial role in pronunciation development.

Second, the results highlight the importance of motivation and active engagement in pronunciation learning. Learners' desire to improve their pronunciation and their practical engagement with English, such as listening to and singing English songs, emerged as factors associated with pronunciation performance. Notably,

enjoyment of listening to English songs was uniquely associated with higher suprasegmental scores, while engagement in singing songs was strongly related to segmental performance. These findings point to the pedagogical value of incorporating music-based activities into pronunciation instruction.

Third, learners' awareness of pronunciation features appeared to follow a hierarchical pattern. The students demonstrated greater awareness and higher performance in basic suprasegmental features than in more complex segmental features, reflecting differences in overall proficiency. Rather than suggesting that suprasegmentals require less instructional attention, this pattern highlights the need to distinguish between current performance levels and pedagogical priorities.

Previous research has often emphasized the role of suprasegmental features in intelligibility. For example, Celce-Murcia et al. (1996) argue that pronunciation errors involving suprasegmentals and connected speech are more detrimental to intelligibility than errors involving individual segmental sounds, and Fraser et al. (2001) notes that native-speaker listeners are more sensitive to stress patterns than to individual phonemes. From this perspective, suprasegmental features may serve as a foundational component for intelligible speech and therefore warrant early and continuous instructional focus.

At the same time, the present findings suggest that segmental features, which showed lower levels of awareness and performance, may require more sustained and targeted instructional support. Given the strong associations between segmental performance and active engagement, such as singing English songs, segmental development may benefit from prolonged, practice-oriented approaches rather than brief or isolated instruction. Taken together, these findings suggest that effective pronunciation instruction for Japanese EFL learners should balance early emphasis on suprasegmentals with continued, focused support for segmental accuracy.

Finally, the results revealed a gap between belief and action. Simply acknowledging the importance of pronunciation showed limited associations with both segmental and suprasegmental performance. This suggests that awareness alone may not be sufficient to improve learners' pronunciation, underscoring the need for instructional approaches that promote active engagement and practice.

5 Limitations

The present study has two main limitations. First, although a sufficient number of participants responded to the questionnaire, the number of students whose audio recordings were included in the ASR analysis was relatively limited. In addition, the relatively small ASR analysis sample (N = 40) may limit the statistical stability and generalizability of the correlation results. Although several significant correlations were identified, the findings should be interpreted cautiously, as small sample sizes may produce unstable estimates of correlation coefficients. Future research with larger speech datasets would allow more robust statistical examination of the relationships between pronunciation awareness and performance, including the reporting of confidence intervals for correlation estimates.

Second, while this study employed an automatic speech recognition (ASR) system to analyze learners' read-aloud speech and previous research reported acceptable reliability of ASR-based pronunciation assessment (*e.g.*, Anguera et al., 2023), ASR systems are not without limitations. Previous studies have shown that ASR output does not always fully correspond to human judgments of pronunciation errors. For example, Nakanishi et al. (2020) compared ASR-generated transcriptions with pronunciation errors identified by multiple trained teachers and reported discrepancies between ASR outputs and human evaluations. Such findings suggest that ASR-based analyses may not capture all aspects of pronunciation accuracy in language learning contexts.

At the same time, recent research has demonstrated substantial improvements in ASR accuracy (*e.g.*, McCrocklin et al., 2019). Moreover, ASR technology offers important pedagogical advantages, including immediate and comprehensive feedback, which can significantly reduce teachers' workload. For learners, ASR-based tools provide opportunities for frequent practice and individualized feedback, which may support the development of speech intelligibility. Therefore, despite its limitations, the use of ASR as both an analytical tool and a pedagogical resource holds considerable potential for pronunciation instruction and second language development.

6 Conclusion

This study investigated the relationships between Japanese EFL learners' attitudes toward pronunciation, pronunciation awareness, and ASR-based pronunciation performance. The findings revealed distinct patterns for segmental and suprasegmental pronunciation, with segmental performance strongly associated with learners' active motivation, confidence, awareness of specific segmental features, and engagement in singing English songs, and suprasegmental performance closely linked to awareness of core prosodic features and enjoyment of listening to English songs. These results suggest that pronunciation development is associated not only with learners' beliefs but also with active, embodied engagement with spoken language. From a pedagogical perspective, effective pronunciation instruction should combine early and continuous attention to suprasegmental features with sustained, practice-oriented support for segmental accuracy. Despite certain methodological limitations, this study highlights the potential of ASR-based analysis as a valuable tool for examining pronunciation development in EFL contexts.

Acknowledgments

The author gratefully acknowledges the cooperation of her colleague teachers, Sachiko Aoki and Rika Takeda, in conducting this study.

References

Anguera, X., Proença, J., Gulordava, K., Tarján, B., Parslow, N., Dobrovolskii, V., & Girard, R. (2023). ELSA Speech Analyzer: English Communication Assessment of Spontaneous Speech. *Proceedings of the 9th Workshop on Speech and Language Technology in Education (SLaTE)* (pp.95-96). Dublin, Ireland.

Armand, S., & Tanaka, Y. (2001). *Japanese Learners' Attitudes toward English Pronunciation* (Doctoral dissertation, Niigata Seiryo University).

Celce-Murcia, M., Brinton, D. M., & Goodwin, J. M. (1996). *Teaching pronunciation: A reference for teachers of English to speakers of other languages*. Cambridge University Press.

Cook, A., & Mishima, A. (2011). *Sounds Like American: A Guide to Fluency in Spoken English*. Cengage Learning.

Ehrlich, S., & Avery, P. (2013). *Teaching American English Pronunciation*. Oxford University Press.

Fraser, H., & Department of Education, Training and Youth Affairs (DETYA). (2001). *Teaching pronunciation: A handbook for teachers and trainers. Three frameworks for an integrated approach.* DETYA. https://helenfraser.com.au/wp-content/uploads/HF-Handbook.pdf.

Jenkins, J. (2000). *The phonology of English as an international language*. Oxford University Press.

McCrocklin, S., Humaidan, A., & Edalatishams, E. (2019). ASR dictation program accuracy: Have current programs improved? In J. Levis, C. Nagel, & E. Todey (Eds.), *Proceedings of the 10th pronunciation in second language learning and teaching conference*. (pp.191-200). Ames, IA: Iowa State University.

Nakanishi, N., Tam, S. Y., & Ebihara, Y. (2020). Spotting English pronunciation errors comparison among teachers and with automatic speech recognition. *LET Kansai Chapter Collected Papers, 18*, 125-146. https://www.jstage.jst.go.jp/article/letkansai/18/0/18_125/_pdf

Ohashi, K. (2018). Research on Japanese University Students' Knowledge, Learning Experience and Attitude Towards English Pronunciation. *JACET Kansai Journal, 20*, 174-193.

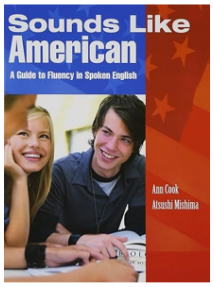
Appendix A

Questions from the Questionnaire

1. Have you learned English conversation?
2. Do you like English?
3. I think English pronunciation is important.
4. I want to improve my English pronunciation.
5. I can pay attention to linking (connecting words) when speaking.
6. I can pay attention to the intonation of sentences when speaking.
7. I can pay attention to where to pause when speaking.
8. I can pay attention to the distinctions between content words (nouns, verbs, adjectives, adverbs, etc.) and function words (prepositions, articles, etc.) while speaking.
9. I can pay attention to word stress (emphasis) when I speak.
10. I can pronounce "r" correctly when speaking.
11. I can pronounce "l" correctly when speaking.
12. I can pronounce "th" correctly when speaking.
13. I can correctly pronounce English consonants that are not mentioned in Questions 9-11.
14. I can correctly pronounce English vowels that do not exist in Japanese.
15. I can pronounce English in an intelligible manner.
16. I listen to English songs.
17. I sing English songs.
18. I like English songs.

Appendix B

Reading Aloud Task for the Students to Record



(p.16)

Passage

Hello, my name is Ann. I'm taking American Accent Training. There's a lot to learn, but I hope to make it as enjoyable as possible. I should pick up on the American intonation pattern pretty easily, although the only way to get it is to practice all of the time. I use the up and down, or peaks and valleys, intonation more than I used to. I've been paying attention to pitch, too. It's like walking down a staircase. I've been talking to a lot of Americans lately, and they tell me that I'm easier to understand. Anyway, I could go on and on, but the important thing is to listen well and sound good. Well, what do you think? Do I?

Teaching Approaches for Enhancing Students' Productive Skills — Focuses on Junior and Senior high school EFL Teachers—

Rise Nagahama[1], Mako Nakada[2], Kokona Nakasone[3], Yui Sugiyama[4], Mai Suzuki[5], Sena Tomoda[6], Yu Tomiyama[7], Ai Yamaguchi[8]

[1-8] **Undergraduate students at Sagami Women's University**

yuya.sagami2312@gmail.com

Abstract

This study aims to clarify how Japanese secondary EFL teachers promote students' productive skills in their EFL classes. In Japanese secondary-level educational contexts, EFL teachers are attempting to improve students' productive skills, aligning with the Japanese national curriculum guideline. In traditional secondary EFL lessons, however, fostering students' productive skills had not sufficiently provided in classroom practices. This research involved 16 secondary EFL teachers all of whom are native Japanese speakers with holding national teacher license. The data was collected through 20 minutes semi-structured interviews. The analysis was conducted using text mining software "KH Coder." The results show that research participants employ collaborative and structured teaching and learning approaches to foster students' productive skills. In speaking activities, pair work and discussion-based tasks were commonly conducted to promote students' interaction and opinion expression. In writing activities, participants focused on sentence-level accuracy and organization. At the same time, participants reported challenges related to limited class hours and the evaluation of productive skills.

Keywords

productive skills, approaches to teaching and learning, challenges, secondary EFL classes

1 Introduction

This study examines how Japanese junior and senior high school EFL supports the development of students' productive skills. Recently, the Fourth Basic Plan for the Promotion of Education (2023–2027) has released that at least 50% of students reach CEFR A1 by the end of junior high school, and that, by the end of senior high school, 60% reach A2 or higher and 30% reach B1 or higher. Despite the above expectations, Sugita (2021) and Uenishi and Yamamoto (2024) point out that classroom practices in secondary schools have not fully achieved the intended development of productive skills Sugita (2021) conducted a survey of 794 university students who had completed three years of EFL lessons in senior high school, found that 458 students identified speaking as their weakest skill, and this is linked to the limited opportunities for oral

production at high school. Uenishi and Yamamoto (2024) report that EFL teachers struggle to design lessons and assessments that effectively foster productive skills. Research on classroom practices for how EFL teachers foster students' productive skills and the challenges they face remain limited. Therefore, examining EFL teachers' teaching and learning approaches are fostering productive skills and teachers' difficulties is essential for clarifying the current situation of EFL.

2 Literature review

At the secondary level, opportunities to develop productive skills have not been sufficiently provided. For instance, the results of the FY2023 National Assessment of Academic Ability and Learning Situation indicate that the implementation rate of performance-based assessments for productive skills remained low: 24.1% for writing, 14.5% for speaking (interaction), and 4.2% for speaking (presentations). Since 2017—when the revised Japanese national curriculum was officially announced—the number of research papers on practices for developing productive skills has increased. For example, CiNii Research related to productive skills between 2010 and 2017 (using search terms such as "English", "speaking", and "writing") yields 120 items over the seven-year period. In contrast, the seven years from 2018 to 2025, the number rises to 287, confirming that the volume of published work has more than doubled. As practical reports published after 2017, studies such as Otsuka et al. (2022) and Sako (2023) provide concrete examples. Otsuka et al. (2022) reports that structured small talk over a 15-week period significantly improved students' speaking ability. Sako's (2023) study suggests that debate-style activities incorporating the AREA model (Assertion, Reasoning, Example/Evidence, Assertion) may improve EFL learners' ability to write and speak. Despite the growing number of practical reports and studies, there is still very limited research that examines how individual secondary school EFL teachers implement practices to foster students' productive skills, and the dilemmas they experience in practice. From this background, this study aims to examine the realities of how productive skills are fostered and to identify the kinds of conflicts teachers face.

3 Methods

3.1 Research Questions and Research Methods

To address the aims mentioned above, the following research questions are established: **RQ1:** What teaching and learning approaches do EFL teachers use to foster students' productive skills? **RQ2:** What kinds of conflicts or difficulties do individual EFL teachers experience when conducting lessons to promote students' productive skills?

The 20 minutes semi-structured interview was conducted in Japanese by the authors with the six questions show in Table 1. The participants are presented in Table 2. All of them are native Japanese speakers with holding Japanese national teacher licence. The interview was conducted online from mid-June to early August 2025. Participants include seven senior high school teachers, four junior high school teachers. Among

them, three had been teaching for less than five years (B, J, and K), three have between five and ten years of experience (E, H and I), and four had been teaching for more than ten years (C, D, F, and G). The interviewees had different years of experience, and they all follow the Japanese national curriculum guidelines. To explore differences in practices and conflicts across levels of teaching experience, we deliberately avoided setting any restrictions on years of experience when recruiting interviewees (Sagami Women's University, Research Ethics Committee, approval number: 25023）.

Table 1

Questions displayed in a semi-structured interview

Q1	What is the final goal of the academic year's speaking (presentations, interaction)?
Q2	What kind of approaches to teaching are practiced for fostering students' speaking skill?
Q3	What are the difficult challenges in teaching speaking?
Q4	What is writing final goals of the academic year?
Q5	What kind of approaches to teaching are practiced for fostering students' writing skills?
Q6	What are the difficult challenges in teaching writing?

Table 2

Participants Information

Teachers	School Types	Teaching experience	Teachers	School Types	Teaching Experience
A	senior high (private)	over 10 years	G	senior high (private)	over 10 years
B	junior high (private)	1-3 years	H	senior high (public)	around 5 years
C	junior high (private)	over 10 years	I	senior high (public)	around 5 years
D	senior high (private)	over 10 years	J	senior high (private)	1-3 years
E	junior high (private)	around 5 years	K	junior high (public)	1-3 years
F	senior high (private)	over 10 years			

3.2 Procedure

KH Coder (Version 3, classroom edition) was used to extract and analyze words from the interview transcripts (16,348 words are extracted). KH Coder is a automated text analysis software, offering functions such as co-occurrence network analysis, and correspondence analysis. It has been widely used to support more objective analyses of qualitative data in fields such as sociology, psychology, and linguistics (as of 24 November 2025, 8,234 relevant papers were identified). In addition, to examine how words connect with surrounding words, we analyzed the narratives using a function called KWIC concordance, which allows the target word to be checked in its immediate left and right context.

3.3 Explanation of Co-occurrence Network and Correspondence Analysis

Co-occurrence network analysis in KH Coder was used to visualise patterns of word co-occurrence in the interview data. Solid lines indicate stronger co-occurrence relationships, whereas dotted lines indicate weaker relationships, and larger circles indicate more frequent words. In contract, correspondence analysis was used

to examine associations between respondents and the words they used. The distance between points indicates the degree of similarity: respondents located closer together are interpreted as having more similar perspectives, whereas those positioned farther apart are interpreted as having more distinct perspectives. Points near the origin (0, 0) represent more average patterns, while points farther from the origin represent more distinctive patterns. The diagram is divided into four quadrants: upper right (1st quadrant), upper left (2nd quadrant), lower left (3rd quadrant), and lower right (4th quadrant).

4 Results and discussions

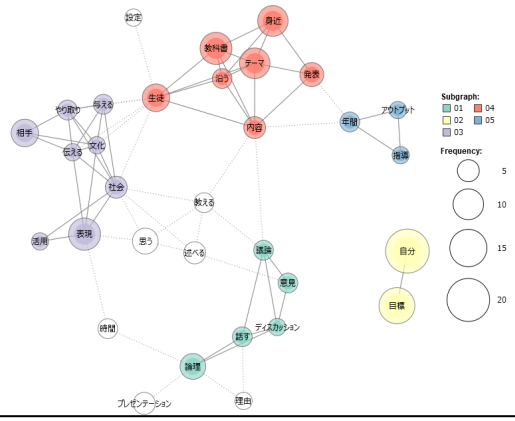
This chapter explores the results of the semi-structured interviews, based on the questions listed in Table 1.

4.1 Results of Question 1 - Co-occurrence Network and Correspondence Analysis

In Figure 1, subgraph 01 indicates that, students are encouraged to engage in discussion with logical opinions and speak actively to develop their speaking skills. Subgraph 02 indicates that the aims of the classes are to enable students to express their own opinions. Subgraph 03 indicates that research participants set goals such as the following: through interactive interaction with a partner, teachers give students opportunities to utilize learned expression so that they can tell and express their opinions about society and culture, while utilizing their knowledge. Subgraph 04 indicates that students are expected to follow the textbook, work on a theme that is familiar to them, and present content based on the textbook materials. Subgraph 05 suggests that the yearly goal is to encourage students to do output supporting with teachers' guidance.

Figure 1

Co-occurrence Network of Q1 and Translation List

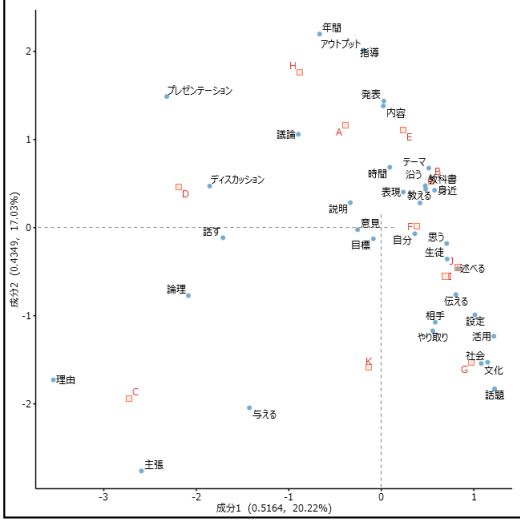


Subgraph 01	logic（論理）, discussion（議論）, opinion（意見）, speak（話す）.
Subgraph 02	self（自分）, aims（目的）.
Subgraph 03	expression（表現), society（社会）, partner（相手）, interaction（やり取り）, give（与える）, utilize（活用）, culture（文化）, and tell（伝える）. Subgraph 03 indicates the words expression（表現), society（社会）, partner（相手）, interaction（やり取り）, give（与える）, utilize（活用）, culture（文化）, tell（伝える）.
Subgraph 04	textbook（教科書）, theme（テーマ）, familiar（身近）, student（生徒）, present（発表）, content（内容）, follow（沿う）
Subgraph 05	output（アウトプット）, yearly（年間で）, and guidance（指導）

In Figure 2, the words of the near origin suggest an overall tendency: the participants set year-end goals that promote self-expression, encouraging students to express and explain their opinions in their own words. In the 1st quadrant, teacher E and H mention the words "output," "present," and "content". They stress that an important goal is to encourage students to produce output by giving presentations on given contents. In the second quadrant, teacher D mentions the words "presentation" and "discussion." Teacher D stresses

Figure 2

Correspondence Analysis of Q1



Japanese-English bilingual comparison

“説明 (explain)”, “意見 (opinion)”, “説明 (explain)”, “自分 (oneself)”, and “目標 (goal)”, “アウトプット (output)”, “発表 (present)”, and “内容 (content)”, “プレゼンテーション (presentation)”, and “ディスカッション (discussion)”, “主張 (claim)”, “理由 (reason)”, and “論理 (logic)”, “伝える (tell)”, “社会 (society)”, “文化 (culture)”, and “相手 (partner)”

the importance of encouraging students to engage in discussions and give presentations in class. In the 3rd quadrant, teacher C mentions the words “claim,” “reason,” and “logic”. This suggests that Teacher C stresses that students should express their own claims with logical reasoning. In the 4th quadrant, teacher G mentions the words “tell,” “society,” “culture,” and “partner.” This highlights that communicative orientation emphasises conveying messages to others within real contexts. Overall, the participants tend to set year-end goals that emphasize students’ ability to express and explain their opinions in their own words, though opportunities to speak through interaction with partners. In addition, classes also tend to follow the textbook content, while using themes that are familiar to students and requiring them to produce output through presentations. The results also suggest that distinctive practices can be mainly grouped into four: 1) an output-oriented approach that focuses on presentations based on given content, 2) combines presentations and discussions to promote active classroom interaction, 3) emphasizes constructing claims with reasons and logical reasoning, 4) emphasizes conveying messages to partners in relation to society and culture.

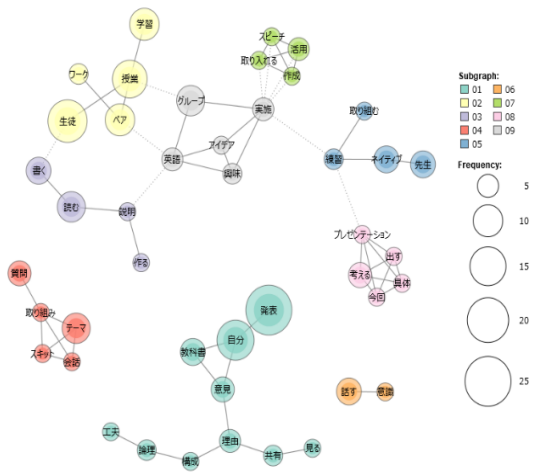
4.2 Results of Question 2 - Co-occurrence Network and Correspondence Analysis

In Figure 3, subgraph 01 indicates that giving a presentation based on the textbook contents encourages students’ self-expression, in which students organize opinions with logical reasoning. Subgraph 02 indicates that participants engage students in pair activities. Subgraph 03 indicates that write, read, explain, and make activities are integrated to support their speaking skills. Subgraph 04 indicates that question-based activities related to a theme are conducted through conversation-focused activities such as skit performances. Subgraph 05 indicates that students work on language practice under the guidance of a native teacher. Subgraph 06 indicates an emphasis on intentional and reflective spoken communication. Subgraph 07 suggests that students incorporate learned content into speech, utilize it actively, and create their own speech script. Subgraph 08 indicates that on presentation activities, participants require students to think carefully and give concrete ideas

by expressing specific and well-developed contents. Subgraph 09 suggests that group-based practices in which students use English to share ideas, implement activities, and sustain interest.

Figure 3

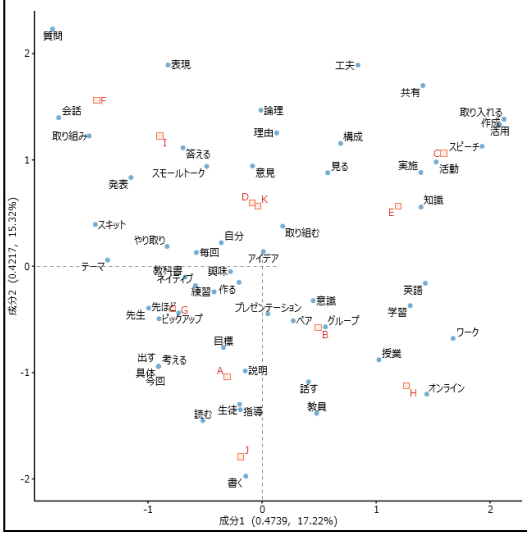
Co-occurrence Network of Q 2 and Translation List



Subgraph 01　presentation,（プレゼンテーション） self（自分）, textbook（教科書）, present（発表する）, opinions（意見）, reasons（理由）, sharing（共有）, look（見る）, composition（構成）, logic（論理）,　devise（工夫）
Subgraph 02　lesson（授業）, learning（学習）, students（生徒）, work（ワーク）, and pair（ペア）
Subgraph 03　write（書く）, read（読む）, explain（説明する）, and make（作る）
Subgraph 04　question（質問）, theme（テーマ）, conversation（会話）, attempt（取り組み）, skit（スキット）
Subgraph 05　work on（取り組む）, practice（練習する）, native（ネイティブ）, teacher（教員）
Subgraph 06　talk（話す）, consciousness（意識）
Subgraph 07　speech（スピーチ）, incorporate（取り入れる）, utilize（活用）, create（作成）
Subgraph 08　presentation（プレゼンテーション）, think（考える）, give（出す）, concrete（具体的）, this time（今回）
Subgraph09　group（グループ）, English（英語）, idea（アイデア）, implement（実施する）, interest（興味）

Figure 4

Correspondence Analysis of Q2



Japanese-English bilingual comparison

“アイデア (ideas)”, “プレゼンテーション (presentations)”, “ペア (pairs)”, “グループ (groups)”, “興味 (interest)”, “練習 (practices)”, “取り入れる (incorporate)”, “作成 (create)”, “活用 (utilize)”, “スピーチ (speech)”, “活動 (activity)”, “表現 (express)”, “質問 (question)”, “発表 (present)”, “スモールトーク (small talk)”, “答える (answer)”, “学習 (learning)”, “ワーク (work)”, “授

In Figure 4, near the words of origin, these suggests that participants encourage students to share and express their opinions and ideas through pair or group and presentation activities based on their interests. In contrast, the words “incorporate,” “create,” “utilize,” “speech,” and “activity” are indicated in the 1st quadrant. This suggests that Teacher C and E encourage students to create their own questions and share them with peers, and that they engage in speeches by making use of mini-games and ChatGPT. In the 2nd quadrant, teacher F and I mention “express,” “question,” “present,” “small talk,” “answer.” This suggests that by providing students with opportunities for output through small talk and presentations, participants encourage them to expand their range of expressions when answering questions or expressing their own opinions. In the 3rd quadrant, teacher

J mentions “read,” “goal,” and “write.” It is suggested that this encourages participants to write down or share their thoughts after students read textbooks or books, thereby fostering the ability to speak logically with supporting reasons. In the 4th quadrant, Teachers B and H mentions “learning”, “work”, “class”, and “online”. This suggests that group / pair work is used to foster students’ ability to summarize information and communicate with others. In addition, online group learning was implemented; enables flexible learning beyond temporal and spatial limitations.

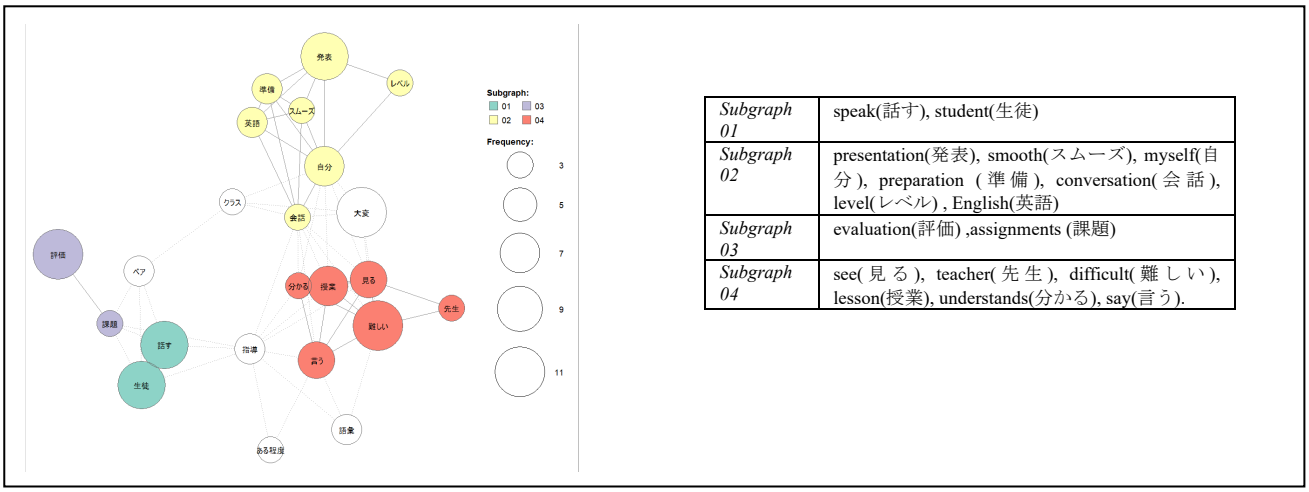
Overall, the results indicate that speaking practices are organized by a combination of structured presentation-based activities and interactive learning. Textbook-based presentations encourage students to organize their opinions logically with reasons, while pair / group work foster active participation, idea sharing, and reflection.

4.3 Results of Question 3

In Figure 5, Subgraph 01 indicates that providing speaking experience, creating opportunities for students to speak, and fostering a willingness to speak. Subgraph 02 indicates that whole-class interactions tend to involve high psychological hurdles for students, as students are expected to prepare in a way that allows them to speak smooth, express themselves, and engage in conversation in English. Subgraph 03 indicates that teachers use different grading methods and that comprehensive evaluation is difficult due to a lack of time and the large number of students. Subgraph 04 indicates it is difficult for teachers to see whether all students understand the task and have acquired the intended skills during pair work. Secondly, students often fail to maintain conversational exchanges in lessons. Due to their limited vocabulary, it is difficult for them to move to a second turn and say what they want to say using appropriate words.

Figure 5

Co-occurrence network of Q3 and Translation List

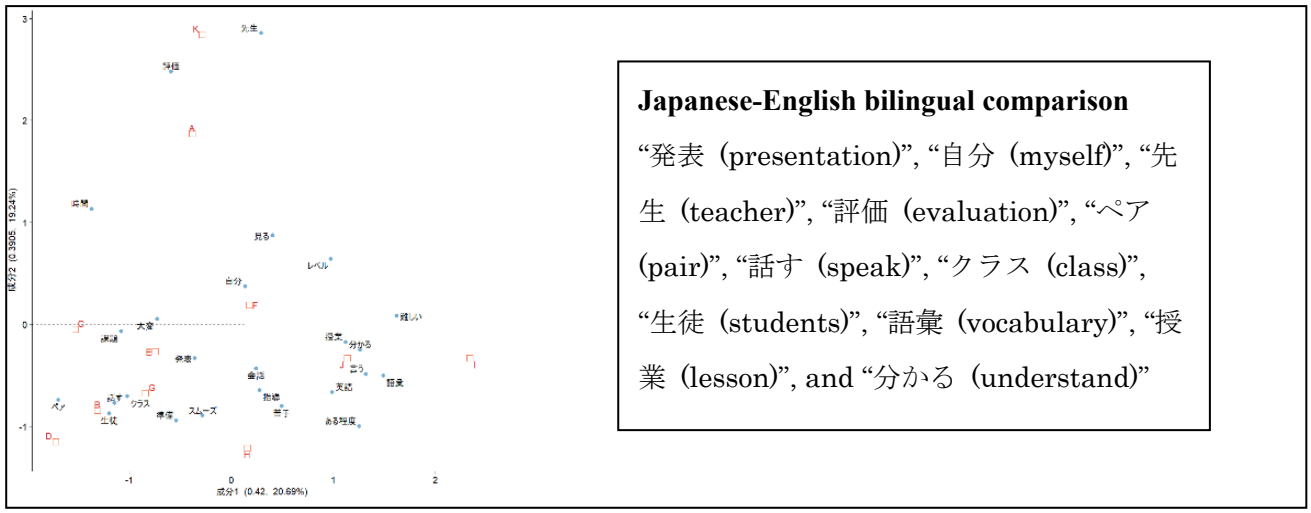


Subgraph 01	speak(話す), student(生徒)
Subgraph 02	presentation(発表), smooth(スムーズ), myself(自分), preparation (準備), conversation(会話), level(レベル) , English(英語)
Subgraph 03	evaluation(評価) ,assignments (課題)
Subgraph 04	see(見る), teacher(先生), difficult(難しい), lesson(授業), understands(分かる), say(言う).

In Figure 6, near the origin, words such as, “presentation, “ and “myself”. This suggests that it is difficult to encourage students to think independently and make claims based on evidence in speaking, and presentation.

Figure 6

Correspondence analysis of Q3



Distinctive features were observed in the following two quadrants. In the 2nd quadrant Teacher A and K suggested that the words "teacher” and “evaluation”. This suggests that Teacher A and K feel difficulty to unify evaluation methods, as the interpretation of evaluation methods and evaluation criteria differs among teachers. In the 3rd quadrant Teachers B and G referred to the words “pair,” “speak, “ “class” and “students,” This indicates that students tend to produce fewer utterances during pair work and struggle to express their ideas in front of a large audience. In the 4th quadrant Teachers I and J referred to the words “vocabulary,” “lesson” and “understand.” This represents that difficulties in teaching how to select and combine appropriate vocabulary. Overall, the results classified into three types: 1) difficulty of encouraging students to think independently and make claims based on evidence, 2) the difficulty of unifying evaluation methods among teachers, the difficulty of students speaking English, and 3) the difficulty of teaching students how to utilize appropriate vocabulary.

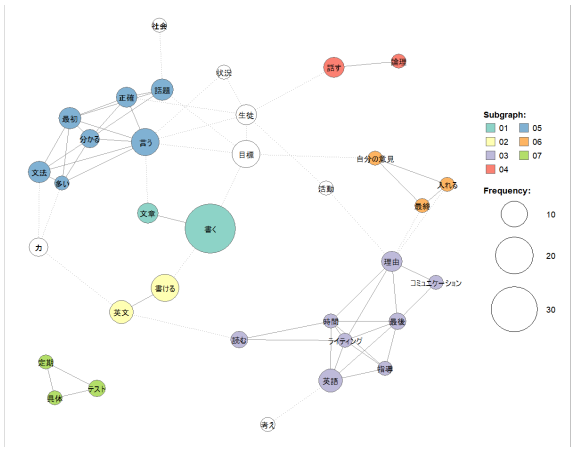
4.4 Results of Question 4

In Figure 7, subgraph 01 and 02 indicate that students develop their ability to write accurately from single sentences to full texts by using familiar topics, paraphrasing, and grammatical understanding. Subgraph 03 indicates that in teaching writing and speaking, emphasis is placed on developing students’ thinking skills from multiple perspectives through reading in English, spending time on writing, considering reasons, and organizing ideas from the beginning to the last, in order to enhance communication. Subgraph 04 and 05indicates that ethical topics are picked up at speaking and the importance of helping students read and understand many English texts and accurately express their own opinions, reasons, and supporting evidence with correct grammar in English on daily and social topics. Subgraph 06 indicates that students aim to put in texts that incorporate their own opinions finally. Subgraph 07 shows that teachers ask follow-up questions

based on students' written texts from regular tests to encourage them to write more concrete content and deepen their ideas.

Figure 7

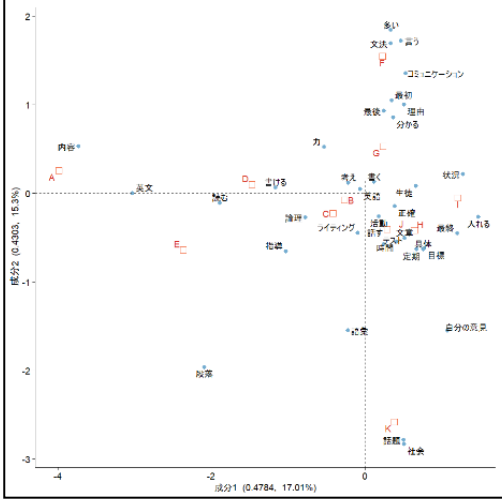
Co-occurrence network of Q 4 and Translation List



Subgraph 01	write(書く) , sentence(文章).
Subgraph 02	English sentences (英文), can write(書ける).
Subgraph 03	read(読む), time(時間), writing(ライティング), English(英 語), reason(理 由), last(最 後), teaching(指導), communication(コミュニケーション)
Subgraph 04	speak(話す),and ethics(倫理).
Subgraph 05	topic(話 題), grammar(文 法), read(読 む), English(英語), accurate(正確), many(多い) , understand(分かる)
Subgraph 06	own opinion(自分の意見), put in(入れる), final(最終).
Subgraph 07	regular(定期), test(テスト), concrete(具体).

Figure 8

Correspondence analysis of Q4



Japanese-English bilingual comparison

"考え（idea)", "書く（write)", "生徒（student)", "英語（English)", "正確（accurate)", "内容（contents)", "英文（English)", "社会（society)", "話題（topic)", "内容（content)", and "英文

In figure 8, near the origin, words such as, "idea", "write", "student"," English" and "accurate" develop students' ability to write their own opinions accurately. In the 2nd quadrant, the words "contents" and "Englis" are located, while in the 4th quadrant "society" and "topic" they are positioned far apart from each other. Distinctive features were observed in the following two quadrants. In the 2nd quadrant Teacher A mentions "content" and "English" improve written expressions by raising questions, and write texts based on evidence. Overall, types are classified into two types; 1) emphasizes evidence and accuracy in writing, 2) focuses on improving expressive skills by using easier and more familiar topics. Ther results indicate that final writing goals for the academic year fall into two tendencies: 1) using daily, easy-to-write content with reasoning, both aiming to improve expressive ability. 2) differences were observed in students' proficiency levels, when

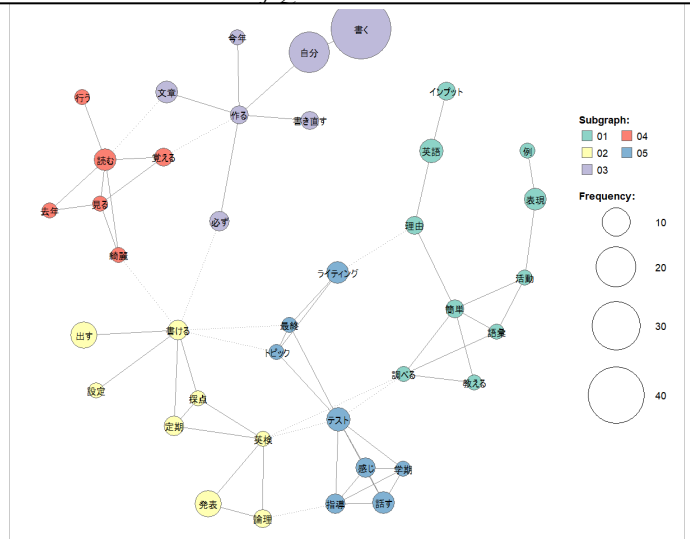
comparing Teacher A and Teacher K, their students' proficiency levels were contrasting, resulting in different prerequisites.

4.5 Results of Question 5

In Figure 9, subgraph 01 indicates that they are being taught to use simple vocabulary and structures when expressing themselves. Some teachers have a rule that students can look up words they don't understand. Subgraph 02 suggests that teachers set a topic and provide opportunities for students to present what they learned in that unit. Teachers also encouraged students to write original content rather than simply writing what is in the textbook. Subgraph 03 indicates that students have an opportunity to express themselves by writing thoughts and feelings. Subgraph 04 indicates that teachers help students practice writing what they learned in class in clear sentences, and then repeatedly revise the sentences so that they can write them on tests. Subgraph 05 shows that they try to do two or three writing activities per semester, and students are sometimes asked to try writing university entrance exam questions.

Figure 9

Co-occurrence Network of Q5 and Translation List



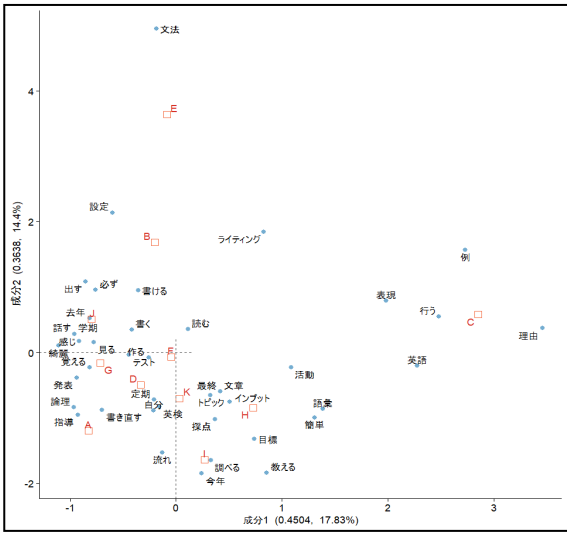
Subgraph 01	Input（インプット）,English(英語), reason（理由）, easy（簡単）, explore（調べる）, vocabularies（語彙）, activities（活動）, expressions（表現）, examples（例）
Subgraph 02	can write（書ける）, submit（出す）, settings（設定）,regular（定期）, grading（採点）, Eiken（英検）, presentation（発表）, logic（論理）.
Subgraph 03	write(書く), rewrite(書き直す), make(作る), documents（文章）
Subgraph 04	read（読む）, memorize（覚える）, do（行う）, see（見る）, last year（去年）, clean(綺麗).
Subgraph 05	test（テスト）, feeling（感じ）, semester（学期）, instruction（指導）, speak（話す）, topics（トピック）, final（最終）, writing（ライティング）.

In Figure 10, words near the origin indicate that teachers foster evidence-based, logical writing skills by having students write about everyday topics and simple situations. In 1st quadrant, Teacher C mentions "example", "expression" and "do". Teacher C emphasis on showing an example of how to use a certain expression. In 2nd quadrant, Teacher B focuses on "setting" of the purpose, scene, and situation. This suggests that students find it easier to write by setting them in addition to example sentences. In the 3rd quadrant, Teacher A, D and G mention "presentation", "oneself", "make". They provide students with opportunities to compose and present texts in their own words using textbook contents. Also, teachers create an environment where students can learn from each other through active reading and presentation opportunities. In the 4th quadrant,

Teacher H mentions “topics”, “input” and “easy”. This suggests the importance of a variety of topics and getting a variety of information. Also, teacher H encourages students to express even simple words in their own words. The results suggest that students are given repeated opportunities to write in conjunction with class activities and assessments, and necessary support in expressing their ideas through topic setting with simple vocabulary and structures, and feedback and rewriting. Also, the results show that teachers provide some example context and example sentences, while others place more emphasis on students expressing their own ideas through presentation activities.

Figure 10

Correspondence analysis of Q5



Japanese-English bilingual comparison

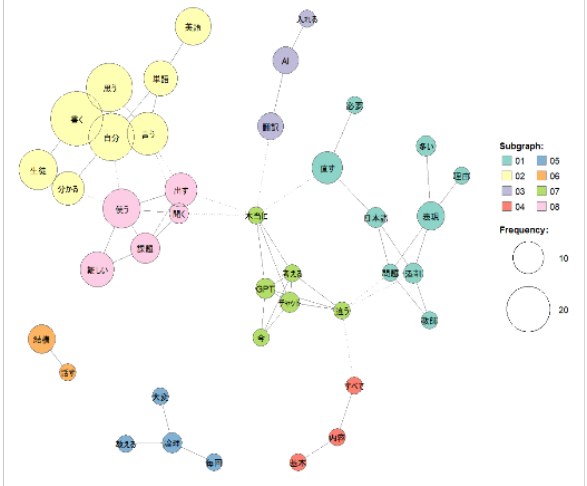
“読む（read)”, “作る（make)”, “テスト（test)”, “書く（write)”, “例（example)”, “表現（expression)”, “行う（do)”, “設定（setting)”, “発表（presentation)”, “自分（oneself)”, “作る（make)”, “トピック（topics)”, “インプット（input)”, and “簡単（easy)”

4.6 Results of Question 6

In Figure 11, subgraph 01 indicate that participants feel there were issues with feedback. Although it is

Figure 11

Co-occurrence Network of Q6 and Translation List

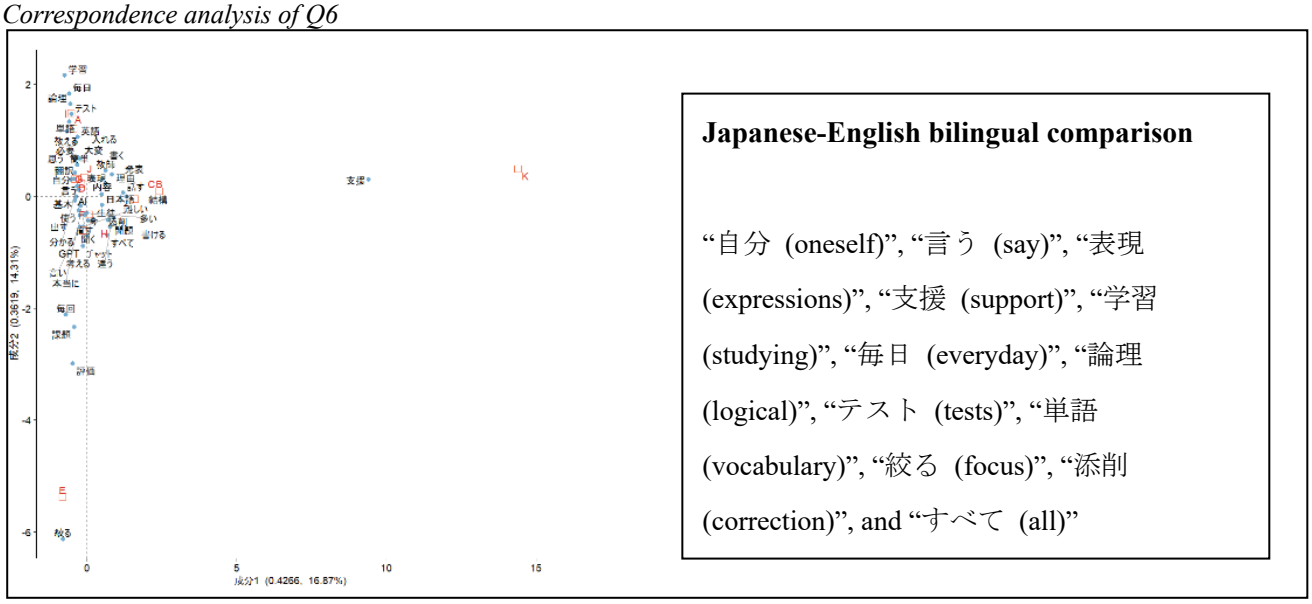


Subgraph 01	English(英語), words(単語), say(言う), think(思う), oneself(自分), write(書く),students (生徒) , underst, (分かる)
Subgraph 02	necessary(必要), correct (直す), Japanese(日本語), problems (問題), teachers(教 師), correction(添 削), expression(表 現), reason(理由) , many(多い)
Subgraph 03	translation(翻訳), AI , input(インプット)
Subgraph 04	translation(翻訳), AI , input(インプット)
Subgraph 05	hard(大変), logic(論理), teach(教える) , every time(毎回)
Subgraph 06	quite (結構) , speak (話す).
Subgraph 07	really(本当に), think(考える), different(違う), chat(チャット), GPT , now(今)
Subgraph 08	use (使う), submit (出す), ask/listen (聞く), assignments (課題), , difficult(難しい).

necessary to correct sentences and expressions, correcting everything will lower students' motivation. Subgraph 02 shows that teachers address problems of students' inability to express reasoning and the unnaturalness of their English expressions. If teachers correct all the students' wiring, it will be too much, so they narrow it down and have them correct it or translate it into Japanese. Subgraph 03 suggests that some students simply input prompts generative AI. Subgraph 04 suggests that difficulties for teachers correct all contents, so basic grading criteria are standardized. Subgraph 05 and 06 suggests that making balance between logic and English level with some difficulties. Also, it is hard to find time to provide feedback. Subgraph 07 indicates that students submit written work using tools such as ChatGPT, and teachers sometimes question whether they can really understand it. Subgraph 08 indicates that even if teachers give some assignments but are strict in giving feedback.

Figure 12

Correspondence analysis of Q6



Japanese-English bilingual comparison

"自分 (oneself)", "言う (say)", "表現 (expressions)", "支援 (support)", "学習 (studying)", "毎日 (everyday)", "論理 (logical)", "テスト (tests)", "単語 (vocabulary)", "絞る (focus)", "添削 (correction)", and "すべて (all)"

In Figure12, the words "AI", "oneself", "say", and "expressions" are indicated in 0,0. They mention that some students use translation machines but end up copying the text that comes out. In the 1st quadrant, Teacher K mentions "support". Teacher K believes that students can write with support but are unsure how much support they should provide. In the 2nd quadrant, Teacher A and I have the same difficulties. They focus on the words "studying", "everyday", "logical", "tests" and "vocabulary". They have introduced a vocabulary learning app and are also conducting online vocabulary tests. However, students sometimes use difficult words directly through translation, so it has become necessary to tell them not to use words they do not understand. It is difficult to teach while maintaining a balance between English ability and logical reasoning. In the third quadrant, Teacher E has a different opinion. It concentrates on "focus". Teacher E mentions that the importance of giving feedback focusing on either structure or grammar. In addition, it was found that it was difficult to provide comprehensive feedback, such as selecting only one sentence from the incorrect sentence and explaining it. In fourth quadrant, Teacher C and F focus on the words "correction" and "all". They commented

that it was difficult to correct all the students' writing. Also, they indicate that simply correcting students' corrections in red ink is not enough and that it is necessary to talk to students directly, which shows that the lack of time for corrections is also an issue.

5 Conclusion and Limitations

Regarding RQ1, the findings indicate that participants employ a range of collaborative and integrated approaches. About speaking, practices such as pair or group work, discussions, and presentations were commonly reported, suggesting that opportunities for interaction and opinion expression are intentionally incorporated into classroom activities. The results suggest that speaking activities are not treated as isolated tasks but are embedded within collaborative learning formats that encourage students to articulate their ideas and respond to others. Regarding writing, the result clarifies a different pattern in teaching and learning approaches. Teachers tended to focus on supporting students at the sentence level. For instance, classroom practices such as reading model texts, rewriting sentences, and providing vocabulary or grammar support were used to scaffold students' writing processes. These findings indicate that writing teaching and learning approaches place particular emphasis on accuracy and organization, and that teachers consciously structure learning opportunities to help students develop written output step by step.

With respect to RQ2, the analysis highlights several challenges that teachers face when promoting students' productive skills. The results indicate that teachers struggle to secure sufficient class time, sustain students' oral interaction, and provide individualized feedback, particularly in writing activities. In addition, teachers expressed concerns about the subjectivity of evaluating productive skills and the increasing complexity of assessment in contexts where students' work may be influenced by generative AI. Importantly, these challenges were observed across teachers with different years of teaching experience. Further research, however, is needed that classroom observations should be conducted in future studies to examine the alignment between participants' narratives and their actual classroom practices.

References

Otsuka, A., Imai, Y., & Wakamoto, N. (2022). "Small Talk" as a tool for improving speaking ability by using a flipped classroom teaching method]. *Studies in English Language Teaching, 45,* 77–94.

Sako, T. (2023). Seito no hihanteki shiko o kojo saseru jugyo-zukuri o mezashite [Aiming to create lessons that improve students' critical thinking]. *KELES Journal, 8*, 68–73.

Sugita, Y. (2022). Improving Speaking Skills in High School English Classes: Insights from Questionnaire Response. *JACET-KANTO Journal, 9*, 5–25.

Uenishi, H., & Yamamoto, N. (2024). Design the lesson to cultivate the students' abilities to think, make decisions and express oneself. Graduate School of Teacher Education Wakayama University Bulletin of course Specializing in Professional Development in Education, 9, 35–40.

Classroom-based Assessment: Beliefs vs. Practice

Taiko Tsuchihira[1], Adam Murray[2], Kei Miyazaki[3] and Yasuko Ito[4]

[1]SEITOKU University, [2]University of the RYUKYUS, [3]TOKAI University, [4]KANDA University of International Studies

ttaiko@wa.seitoku.ac.jp, murray@cs.u-ryukyu.ac.jp,
keimiyazaki67@gmail.com, yasukoi@kanda.kuis.ac.jp

Abstract

This study explores the gap between Japanese EFL teachers' beliefs and their actual classroom assessment practices. It examines the relationship between teachers' beliefs, language assessment literacy (LAL), and their confidence in implementing assessment. An explanatory sequential mixed methods approach consisting of two phases was used. In the first phase, the Classroom-based Assessment Self-Efficacy Scale (CBA-SES), which includes LAL items, was given to 50 EFL teachers in Japan. Nine participated in detailed follow-up interviews to better understand their perceptions of CBA. In the second phase, five secondary school teachers were interviewed to explore specific contextual barriers and differences. Results showed a notable gap: teachers firmly believe in the importance of CBA but struggle to implement it. Despite high levels of self-efficacy, their classroom activities were often constrained by external factors, including testing pressures and school policies. Teachers also expressed a need for professional growth in CBA. Overall, the study highlights that challenges in applying effective CBA are rooted more in contextual challenges than in a lack of individual knowledge.

Keywords

CBA, beliefs, self-efficacy, practices, language assessment literacy

1 Introduction

In recent years, there has been a growing interest among EFL educators and researchers in Japan about teachers' attitudes towards classroom-based assessment (CBA) practices. Despite the central role of assessment in EFL teaching, there remains limited data on teachers' beliefs, self-efficacy regarding assessment, and their actual classroom implementation, particularly across different school settings. Previous studies suggest that teachers' perceptions and confidence can significantly influence their assessment practices, which, in turn, affect student learning outcomes. In fact, classroom teachers must have high levels of assessment literacy, which is practical

know-how (skills) along with "relevant background in measurement and language description" (Davies, 2008, p. 328). Teacher self-efficacy (TSE) beliefs can be defined as "teachers' beliefs in their abilities to support learning in task-, domain- and context-specific cognitive, metacognitive, affective and social ways" (Wyatt, 2018, p. 93). In the Japanese EFL context, there has also been a growing interest in beliefs and self-efficacy (Nishino, 2012; Wicking, 2017; Thompson & Woodman, 2018; Yada et al., 2019). Therefore, exploring these attitudes within the Japanese context can help inform future professional development efforts and promote more effective assessment practices.

The researchers have been developing an instrument, the Classroom-based Assessment Self-Efficacy Scale (CBA-SES) (Murray & Tsuchihira,2022), comprising four sections: biographical information and teaching context (11 items), teacher beliefs (11 items), self-efficacy (10 items), and teaching practices (12 items). In Phase 1, 50 teachers from various institutions across Japan completed the questionnaire. Follow-up interviews were conducted with nine participants. Most of the Phase 1 participants were university instructors, whose teaching and assessment may be likely to be affected more by their own institutional constraints and less by MEXT's guidelines. This suggests that the picture may look different with secondary school teachers. We, therefore, conducted the second phase to explore the gap between beliefs and practices among secondary school EFL teachers in Japan. Phase 2 employed an exploratory qualitative approach involving in-depth interviews with five secondary school EFL teachers. The aim was to delve into their current CBA assessment conditions, explore the specific gaps between their Language Assessment Literacy (LAL) beliefs and practices, and analyze their needs for future professional development. The interviews focused on identifying the constraints teachers face and aspects of assessment such as practical knowledge, Assessment for Learning (AfL), and assessment theory.

2 Literature Review

2.1 Teaching Context

In recent years, the Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT) has emphasized language output by setting CEFR target levels of A1 and A2 for junior and senior high school graduates. To do this, MEXT expects teachers to integrate assessments with classroom instruction. Specifically, students should be evaluated along three dimensions: knowledge and skills, thinking and expression, and learner engagement (MEXT, 2020). Ongoing formative assessment should be used not only for grading, but for instruction. Performance tasks for skills such as speaking and writing are strongly encouraged. Students should be regularly provided with constructive feedback and opportunities for self-assessment and reflection. Summative assessment is to be criterion-referenced in alignment with the learning objectives described by MEXT's Course of Study.

MEXT's initiatives include the provision of teacher handbooks featuring assessment samples, the

promotion of speaking and writing tests, and a move toward criterion-referenced evaluation (NIER, 2021). Furthermore, the implementation of the National Core Curriculum for Teaching English emphasizes both pre-service and in-service teacher training (MEXT, 2018). This curriculum identifies LAL as a crucial element of effective L2 education, encouraging EFL teachers acquire the knowledge and skills necessary to design diverse assessment tools. At the regional level, MEXT (2020) requires local governments to ensure that these policies are translated into classroom practice, specifically by increasing the frequency of assessments of productive skills.

However, despite these extensive top-down initiatives, a significant gap remains between official policy and the practical realities of Japanese classrooms. Teachers are now required to transform complex student performances into quantifiable grades for report cards. This pressure to provide objective data for accountability often leads teachers to prioritize summative results over formative assessment. As noted by Koizumi (2022), these national frameworks provide only general assessment structures and limited samples of rubrics and tasks. Because the guidelines lack context-specific issues, the burden of assessment falls on individual teachers. Consequently, this results in inconsistencies.

Because the responsibility for designing and implementing specific assessment tools rests largely with individual teachers, teacher self-efficacy is a critical determinant of classroom practice (Waddington, 2023). This relationship between teacher beliefs and practice creates a divergent landscape: teachers with low self-efficacy often feel hesitant about the ambiguity of general MEXT guidelines. They may avoid implementing complex assessment behaviors altogether. In contrast, those with high self-efficacy tend to create their own evaluation criteria. However, these highly confident teachers are also more likely to prioritize their own standards over official frameworks, potentially leading to assessment practices that diverge significantly from national policy goals for standardized assessment.

2.2 Teacher Self-Efficacy Research in the Japanese Context

Within the Japanese educational context, there has been little or no research concerning self-efficacy in LAL. However, there are a few studies on teachers' self-efficacy. These studies report that Japanese teachers have lower self-efficacy than their international counterparts. A study by Shimojo et al. (2023), drawing on international survey data, revealed this concerning trend, particularly highlighting a gap in self-efficacy related to students' affective and motivational outcomes. Shimojo et al. (2023) also reported that the proportion of Japanese elementary and junior high school teachers with high levels of self-efficacy is notably low. The deficit is particularly pronounced in areas related to enhancing students' affective domain and motivation. For example, only 41.4% of Japanese teachers responded positively ("very well" or "quite a bit") to the statement "helping students see the value of learning," compared with an average of 86.3% across participants from 48

countries. This evidence strongly suggests a prevalent pattern of low self-efficacy among Japanese teachers, particularly regarding their ability to positively influence students' motivation to learn (Shimojo et al., 2023).

While the general trend indicates low self-efficacy, research indicates that direct, positive evidence of student engagement and learning can effectively boost teachers' instructional self-efficacy. This implies that self-efficacy is not fixed and can be positively influenced by observable outcomes of teaching practices. Two studies, one focusing on assessment methods and the other on real-time emotional feedback, illustrate this crucial connection. The study by Yato et al. (2015) focused on the impact of portfolio assessment on junior high school teachers. The researchers found that teachers who used portfolio assessment reported increased self-efficacy after observing improvements in student understanding. Specifically, the positive and reflective narratives written by students in their portfolios provided tangible evidence of learning progression, thereby validating the teachers' instructional efforts and strengthening their confidence in the effectiveness of their teaching strategies. This demonstrates a clear link: when teachers can verify the efficacy of their instruction through positive student responses and demonstrable learning gains, their professional self-efficacy is enhanced (Yato et al., 2015). In another study, teacher self-efficacy was significantly higher among teachers who received visualized emotional feedback (Matsubara & Umemuro, 2023). Furthermore, a significant interaction effect was found between the presence of visualization and the content being taught, suggesting that the visualization of student emotions is more effective when the teacher perceives the content as complex. These findings not only support the idea that immediate student feedback can enhance teacher self-efficacy but also offer practical implications for the design of online learning systems to mitigate the challenges associated with virtual teaching environments.

Both Yato et al. (2015) and Matsubara and Umemuro (2023) provide compelling evidence that visible, positive student reactions and evidence of learning are potent sources for improving teachers' instructional self-efficacy. While the general trend of low Japanese teacher self-efficacy (Shimojo et al., 2023) and the positive impact of visible student success on instructional efficacy (Yato et al., 2015; Matsubara & Umemuro, 2023) are established, a significant gap remains in the literature concerning Japanese EFL teachers' self-efficacy specifically within the domain of LAL. It encompasses the knowledge, skills, and beliefs that teachers hold about planning, framing, conducting, and using (Hill & McNamara, 2012), and it takes into account practice, concepts, and context (Fulcher, 2012). Previous studies have established that Japanese teachers face difficulties connecting their instruction to students' affect and motivation, which are inherently linked to the core functions of CBA and AfL.

Despite research establishing low levels of self-efficacy among Japanese teachers and the positive impact of visible student success on instructional efficacy, a significant gap remains regarding Japanese EFL teachers' self-efficacy, particularly in the domain of LAL. Since assessment is a primary means by which teachers gauge the success of instruction and foster student motivation, low LAL self-efficacy may be a major contributor to the broader issues identified in general teacher self-efficacy. This study is necessary to identify

the current assessment conditions and LAL needs of EFL teachers in Japan. By exploring the gaps between their beliefs, self-efficacy, and assessment practices, this study identifies the specific areas requiring professional development, thereby providing practical insights for enhancing teachers' LAL.

3 Study

3.1 Phase 1

For Phase 1 of this study, an explanatory sequential mixed methods design was employed. This approach was taken because the researchers wanted to identify trends in general beliefs, self-efficacy, and practices regarding CBA. The follow-up interviews were used to better understand the reasons for the participants' responses.

3.1.1 Participants

Participants for Phase 1 were recruited in two ways: audience members at conference presentations, and social media posts by the authors. A total of 50 teachers located in Japan participated in the study. In terms of English-language ability, the participants self-reported by answering, "How do you identify yourself as a user of English?" Thirty were native speakers of English, 10 were EFL speakers, eight were ESL speakers, and two did not respond. Most of the respondents (40) had 11 or more years of teaching experience. Most of the participants (42) taught at universities.

3.1.2 Classroom-based Self-Efficacy Scale (CBA-SES)

In this phase, the participants completed an online questionnaire (Google Form). To protect the participants' anonymity, names and emails were not required. Prior to the questionnaire, participants were provided with detailed information about the research purpose, the voluntary nature of participation, and their right to withdraw from the study at any time without penalty. By proceeding to complete the survey, participants provided their informed consent to participate in the study. The Classroom-based Assessment Self-Efficacy Scale (CBA-SES) is a bilingual instrument that the authors have been developing to investigate the state of classroom-based assessment in the Japanese context. The CBA-SES is based on the Michigan Assessment Literacy Standards (2017), Nishino (2012), and Thompson and Woodman (2018). The instrument comprises four sections, with a total of 57 items (see Appendix A for selected items). The first section of the instrument comprises 11 items on participants' biographical information/teaching context, such as English proficiency, qualifications, and courses taught. The following section has 16 statements about beliefs on CBA, such as "The purpose of assessment is to give grades to students." The third section about self-efficacy has statements such as "I can use learning targets aligned to the standards (e.g., MEXT Course of Study) to guide instruction." The final section on classroom practices covers sources for test items and procedures, forms of assessment and feedback, and time spent on assessment.

3.1.3 Follow-up interviews

Because most respondents (42) were university teachers, we focused our attention on them. In the second part of Phase 1, follow-up interviews were conducted with nine participants by Zoom or email. The Zoom interviews were recorded and transcribed using the built-in transcription feature. The semi-structured interview consisted of 10 questions about the participant's specific teaching context, with questions (see Appendix B for selected items) such as "Do you sometimes face difficulties in evaluating students with your institution's English program or curriculum?"

3.1.4 Results and Discussion

The participants were divided into groups based on their agreement with the assessment statements. The maximum score was 80. The average score for all participants was 65 with a standard deviation of 6.84. Cronbach's alphas for each subscale were .64 (Belief), .88 (Efficacy), and .79 (Practice). The teachers were then divided into three groups: low beliefs (5), average beliefs (36), and strong beliefs (9). The highest level of agreement was with the statements "Teachers should understand what tests are and be able to use them" (4.66), "Quality assessments are a critical attribute of effective teaching and learning" (4.62), and "Effective feedback is necessary for learning" (4.55). This indicates that teachers view assessment as an important component of classroom instruction.

Similarly, participants were divided according to their agreement with the efficacy statements. The maximum score was 65. The average score for all participants was 49.9 with a standard deviation of 8.44. This resulted in three groups: low efficacy (8), average efficacy (33), and high efficacy (9). The highest level of confidence was to the statements "I use different instructional methods according to learning targets" (4.38), "I can use assessment results appropriately to modify instruction to improve student achievement" (4.08), and "I can select and use various assessment methods appropriate to assessment purposes and learning targets" (4.00). This indicates that teachers are confident in their ability to implement classroom-based assessment.

Among the nine university teachers who participated in the second stage of Phase 1, most held average beliefs (8), and one held strong beliefs. In terms of efficacy, their scores were: low (3), average (5), and high (1). One of the questions focused on institutional support available (Appendix B #1, 3). Most interviewees reported relying on resources provided by publishers. They also use standardized tests such as TOEIC and TOEFL. Several teachers (4) mentioned customizing the publisher's resources when making final examinations to fit the learning objectives of their courses. Another commented that their institution provided little support and that teachers were free to assess their students as they saw appropriate. Two teachers mentioned that their institutions provide faculty development lectures/workshops, but attendance is low because the training is unsystematic. At one teacher's institution, a Learning Management System (Moodle) and in-house rubrics are used.

The interviewees were asked about the difficulties of assessing their students (Appendix B #6). They talked about a range of problems they face. The most common theme was how to balance summative and formative assessment in their classes. Similarly, they mentioned the use of weighting when calculating final grades. They mentioned the difficulty of making appropriate rubrics for their specific speaking and writing courses. As noted earlier, many teachers rely on publisher resources. One teacher discussed difficulties at his institution because students had obtained the teaching materials, along with all the tests and answer keys. One teacher talked about students enrolled in speaking or presentation courses with special needs, such as social phobia.

They also feel external pressure from university administration. One teacher mentioned that his university has guidelines about grade distribution. For example, only a certain percentage of students can receive an "S" grade. One teacher mentioned the university's attendance policy and the need to distinguish between passive attendance and active participation. Similarly, another commenter noted that at his institution, participation did not constitute a substantial portion of the final grade. Finally, a teacher described a mismatch between the course content (speaking and oral skills) and the external exams used for grading.

Teachers felt that their institutions needed to provide greater support for teachers. Some of their opinions were general in nature. Subscriptions to test-making software, grammar and plagiarism checkers, and automated assessment tools were mentioned by several interviewees. Other teachers discussed the need for shared resources (teaching plans, rubrics, test items) within their institutions. Finally, they seek additional faculty development opportunities in topics such as rubrics, Learning Management Systems (Moodle), test development, and grade calculation. One teacher noted that lists/can-do statements could be helpful for consciousness-raising but might be perceived negatively by part-time teachers.

3.2 Phase 2

Phase 1 focused mainly on university instructors. The picture looks different with secondary school EFL teachers. Thus, Phase 2 focused on secondary school EFL teachers to explore the gaps between their LAL beliefs and classroom practices, and clarify their needs for professional development, employing in-depth qualitative interviews.

3.2.1 Participants

The interviewees included three male and two female teachers, spanning an age range from the 20s to the 50s. Their teaching experience ranged from 1 to 30 years. Interviews were conducted following the guidelines of the Collaborative IRB Training Initiative. Sampling was purposive to capture a diverse range of perspectives relevant to LAL within Japanese secondary schools. Regarding school type, the participants were drawn from two private university-affiliated schools (Teachers A and B), a national university-affiliated school (Teacher

C), a public school (Teacher D), and a private school with college-bound graduates (Teacher E). Most interviews were conducted online, with one conducted face-to-face. The duration of the interviews ranged from 35 minutes to 60 minutes. All participants received an informed consent form clearly explaining the study's purpose, data collection procedures, confidentiality measures, and their right to withdraw from the study at any point. All collected data were anonymized and stored securely. The interview recordings were password-protected, and the transcribed data were de-identified.

3.2.2 Procedures and Analyses

To explore gaps and analyze teachers' needs for future LAL training, we conducted interviews with five secondary school EFL teachers. This exploratory study aims to explore current assessment gaps and the needs for future LAL training. Interview questions focus on the gaps the teachers have identified in assessment and on the needs for future LAL training. The researcher meticulously checked transcriptions, highlighting, bolding, underlining, and color-coding rich or significant quotes and passages. The researcher also concurrently wrote reflections and memos, and assessed participants' attitudes and behaviors, using coding schemes (Saldaña, 2009). The study employed a combination of descriptive and value coding to categorize the contents of inventory, summaries, and data, and to examine what they do and emphasize in real-world contexts (Leavy, 2014). In addition, the values coding emphasizes interviewees' attitudes, beliefs, "motivation, agency, causality or ideology" (Saldaña, 2009, p.90). To enhance reliability, the researcher reported the identified gaps in the interviews to the co-researchers, and emails were sent to participants who had joined interviews to confirm that their responses accurately reflected their intentions.

3.2.3 Results and Discussion

Phase 2 focuses on the discrepancy between secondary school teachers' beliefs and practices. According to Phase 1, teachers share strong agreement in their beliefs, but these beliefs are not always reflected in their self-efficacy or classroom practices. In particular, secondary school teachers reported fewer CBA practices (See Appendix A).

The interview results indicated that teachers consider CBA an important element of language teaching, but that it is challenging to implement. For example, Teacher A (50s) has extensive teaching experience and has not used the teachers' manuals or assessment criteria offered by the MEXT. Instead, he uses his own assessment criteria, including higher-order thinking. His struggle is that the MEXT-authorized textbook is not suitable for assessing higher-order thinking; instead, it is designed around the typical language knowledge, skills, and performance required by external standardized tests. He faces challenges in implementing his assessment method because of the current trend toward higher external test scores. He recognizes the challenges of assessing higher-order thinking skills using textbooks and employs supplementary materials to address them.

Teacher B (40s) appears to value formative assessment and demonstrates strong self-efficacy in designing and implementing assessments, as evidenced by his use of pop quizzes and his emphasis on providing timely feedback. However, there is limited explicit use of formative assessment strategies due to his colleagues' concerns about unfair grading. His English department requires a unified assessment, suggesting that it is difficult for a single teacher to grade students and provide instruction based solely on formative assessment.

Teacher C (30s) discussed the importance of accountability and the need to justify assessment decisions, highlighting his understanding of the importance of validity and reliability in assessment. While he primarily focuses on summative assessment, he acknowledges the importance of formative assessment in providing feedback and monitoring student progress. The school leaves assessment to individual teachers, but teachers are aware of their accountability for their grades. Thus, in most cases, he uses term tests, quizzes, and performance tests as assessment data for accountability, thereby demonstrating a commitment to data-driven decision-making. In this sense, he emphasizes summative assessment rather than formative assessment. He is not confident in the accuracy of the assessment and is keen to further his training. He wants to learn assessment theory (validity and reliability) and AfL.

Teacher D (20s) does not make any tests or assessment criteria. She relies heavily on existing teachers' manuals and on assessment materials prepared by the school. For her, this is comfortable and easy to assign grades to her students. She believes the teachers' manual and the assessment criteria recommended by MEXT are fair and reliable for both students and teachers. Her assessment practices are influenced primarily by the school's policies and guidelines and by her senior colleagues. However, she also recognizes that her limited autonomy in designing assessments may hinder her ability to ensure their validity. If she were asked to assess students in her own way, she feels she would be upset over this situation due to a lack of assessment literacy. In this sense, she demonstrates a willingness to participate in LAL training if an opportunity arises. She wants to learn about feedback, washback, Assessment of Learning (AoL), AfL, and Assessment as Learning (AaL).

Teacher E (40s) shares a similar perspective with Teacher A, expressing satisfaction with the autonomy she has in designing her own assessment methods. While she teaches techniques for university entrance examinations, she is confident in her instruction of the *English Communication* course, where she employs a task-based approach. As a highly proficient English speaker, she guides students to notice the structures of the English language through engaging activities. However, she expresses a lack of confidence in formative assessment and an intention to acquire more knowledge in this area. She uses publisher-provided rubrics for her task-based lessons, yet summative assessment accounts for most of her evaluation. Due to a heavy workload, she finds herself unable to create her own rubrics. Regarding her beliefs, she considers task-based instruction the correct approach; however, because task-based teaching is not synonymous with formative assessment, she lacks confidence in its practical application and has not implemented it in her practice.

The interviews with five secondary school EFL teachers reveal a complex interplay of beliefs, self-efficacy, and practices, highlighting significant gaps in the implementation of CBA. The teachers consistently recognize the importance of CBA, timely feedback, and formative assessment as key elements for effective language instruction. However, their beliefs often clash with the realities of their assessment practices. This divergence is driven mainly by contextual constraints, whereby the school environment constrains CBA through pressures from standardized testing, adherence to existing school and government policies, and the resulting prioritization of summative assessment for accountability. To bridge these gaps – specifically the heavy reliance on summative methods and limited autonomy – teachers emphasize the need for enhanced LAL training focusing on both theoretical aspects (validity, reliability, general assessment theory) and practical applications, particularly AfL and AaL. This suggests that the difficulty in implementing CBA is not solely an issue of individual knowledge but is significantly compounded by institutional factors, material limitations, and systemic pressures that favor quantifiable results over more holistic, formative approaches.

4 Conclusion

Our findings highlight a gap between EFL teachers' strong beliefs in the importance of CBA and their actual practices. While teachers generally acknowledge the value of timely feedback, formative assessment, and diverse assessment methods, their implementation is heavily constrained by contextual and systemic factors. Specifically, the pressure for accountability and the emphasis on standardized external testing compel teachers, particularly those in secondary schools, to prioritize summative assessment and quantitative data over formative assessment. This institutional environment constrains teachers' autonomy in designing assessments, leading to practices that often contradict the national policy goals of integrated instruction and assessment.

This general awareness of the gap between belief, including self-efficacy, and practice prompts a strong demand for enhanced LAL training. Teachers explicitly requested professional development focused on both theoretical concepts (validity, reliability, assessment theory) and practical applications (AfL and AaL).

As for future research, the findings suggest that addressing the difficulty in implementing effective CBA requires moving beyond individual knowledge deficits to tackle contextual constraints. Future research should investigate how institutional structures can be reformed to provide teachers with greater access to AfL and AaL. Furthermore, studies should focus on developing and evaluating LAL training programs that integrate theoretical knowledge with practical strategies to embed formative assessment into daily teaching, thereby building teacher self-efficacy to overcome existing constraints.

References

Davies, A. (2008). Textbook trends in teaching language testing. *Language Testing, 25*(3), 327–347. https://doi.org/10.1177/0265532208090156

Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly, 9*(2), 113–132. https://doi.org/10.1080/15434303.2011.642041

Hill, K., & McNamara, T. (2012). Developing a comprehensive research framework for classroom-based assessment – an empirical study. *Language Testing, 29*(3), 395–420. https://doi.org/10.1177/0265532211428317

Koizumi, R. (2022). L2 speaking assessment in secondary school classrooms in Japan. *Language Assessment Quarterly, 19,* 142–161. https://doi.org/10.1080/15434303.2021.2023542

Leavy, P. (2014). *The Oxford handbook of qualitative research*. Oxford University Press.

Matsubara, R. & Umemuro, H. (2023). Effects of visualization of student emotions on performance and self-efficacy of teachers in online learning. *The Japanese Journal of Ergonomics, 59*(2), 63-72.

Michigan Assessment Consortium. (2017). *Assessment literacy standards*. Michigan Assessment Consortium.

Ministry of Education, Culture, Sports, Science and Technology (MEXT). (2018). *Gaikokugo (Eigo) Core Curriculum [The proposition of Core Curriculum for foreign language (English*)]. https://www.mext.go.jp/b_menu/shingi/chukyo/chukyo3/002/siryo/attach/1388110.htm

Ministry of Education, Culture, Sports, Science and Technology (MEXT). (2020). *Reiwa ni nendo eigo kyoiku kaizen puran [Plan for improving English language education in 2020]*. https://www.mext.go.jp/a_menu/kokusai/gaikokugo/1418086_00001.htm

Murray, A. & Tsuchihira, T. (2022). Japanese teachers' attitudes towards classroom-based assessment. In P. Ferguson, B. Lacy, & R. Derrah (Eds.), *Learning from Students, Educating Teachers – Research and Practice* (pp. 196–206). JALT. https://doi.org/10.37546 /JALTPCP2022-23

National Institute for Educational Policy Research (NIER). (2021). Shido shiryo jirei shu [Handbook of teaching materials and case examples]. https://www.nier.go.jp/kaihatsu/shidousiryou.html

Nishino, T. (2012). Modeling teacher beliefs and practices in context: A multimethod approach. *The Modern Language Journal, 96*(3), 380–399. https://doi.org/10.1111/j.1540-4781.2012.01364.x

Saldana, J. M. (2009). *The coding manual for qualitative researchers* (2nd ed.). SAGE Publications.

Shimojo, M., Kwon, H., Taira, H., & Uehara, K. (2023). Effects on individual optimization of learning and effects on teachers' self-efficacy through classroom practice based on universal design of learning. *Ryukyu University Education Department Journal, 103*, 35–45.

Thompson, G, & Woodman, K. (2018). Exploring Japanese high school English teachers' foreign language teacher efficacy beliefs, *Asia-Pacific Journal of Teacher Education*, https://doi.org/10.1080/1359866X.2018.1498062

Waddington, J. (2023). Self-efficacy. *ELT Journal, 77*(2), 237–240. https://doi.org/10.1093/elt/ccac046

Wicking, P. (2017). The assessment beliefs and practices of English teachers in Japanese universities, *JLTA Journal, 20*, 76–89. https://doi.org/10.20622/jltajournal.20.0_76

Wyatt, M. (2018). Language teachers' self-efficacy beliefs: A review of the literature (2005-2016), *Australian*

Journal of Teacher Education, 43(4), 92–120. https://doi.org/10.14221/ajte.2018v43n4.6

Yada, A., Tolvanen, A., Malinen, O., Imai-Matsumura, K., Shimada, H., Koike, R., & Savolainen, H. (2019). Teachers' self-efficacy and the sources of efficacy: A cross-cultural investigation in Japan and Finland. *Teaching and Teacher Education, 81*, 13–24. https://doi.org/10.1016/j.tate.2019.01.014

Yato, S., Nakajima, M., & Hori, T. (2015). A study on the effective use of OPPA : mainly through the method of teaching relative clauses in high school, and the teachers' feelings of self-efficacy caused by OPPA. *Journal of the Center for Educational Research and Teacher Development, Yamanashi University, 20*, 227–240.

Appendix A

Selected CBA-SES Items

Belief
(B 13) A purpose of assessment is for giving grades to students.
(B 24) Students should learn how to use assessment results to improve their learning.
Self-efficacy
(E 6) I can use grading practices that result in grades that are accurate, consistent, meaningful and supportive of learning.
(E 10) I can make tests that resemble real-life use of English.
Practices
(P 2) I make tests that resemble real-life language use.
(P 7) I use summative assessment.

Appendix B

Selected Follow-up Interview Questions

1. To what extent do you think your institution provides you (the classroom practitioner) with (or expects you to use) the resources for the effective assessment of your students' English language skills? (e.g., textbooks with built-in assessment, online quizzes, teacher-made paper-pencil tests, teacher-made online tests, teacher-made face-to-face speaking tests, presentation tests, external tests (TOEIC, EIKEN), worksheets, online self-study assignments, reaction papers (exit tickets), self-assessment checklists, peer assessment worksheets, journaling, concept (mind) maps, online quizzes, online games)
3. To what extent do the administration (such as the university curriculum committee and textbook publishers) provide assessment resources? (such as teacher guides or manuals, assessment frameworks or rubrics, websites or online platforms, professional development workshops or courses, or books/articles on assessment?
4. What support should your institution provide in order to develop your knowledge and skills of assessment?
6. Do you sometimes face difficulties in evaluating students with your institution's English program or curriculum?

Analyzing Tourism English and Machine Translation: Supporting Tourism in Japan

Yoko Sakamoto[1], Yukiko Ideno[2] and Teppei Terashima[3]

[1]Hakuoh University, [2]Tokiwa University, [3]Tokiwa University

y-saka@fc.hakuoh.ac.jp, y-ideno@tokiwa.ac.jp, terasima@tokiwa.ac.jp

Abstract

This research analyzes the effectiveness of the machine translation (MT) tools to assist foreign tourists who visit Japan and to be applied in the teaching of tourism English. The area of study is the international airports in Haneda and Narita which are the ports of entry of international tourists. The need for multilingual communication has increased since tourism is on the rebound after the COVID-19 pandemic, yet most of the attractions in Japan do not have English-speaking staff and good signage, especially in the rural areas.

To understand communication in everyday life, we observed airport staff and tourists using mobile translation applications such as Google Translate and Pocketalk applications. We also examined English travel phrases in airport signage and service manuals and travel brochures to see how well MT systems translate them. Our research demonstrates that MT tools promote tourists' confidence and independence, motivating them to visit regional areas such as Hokkaido and Fukuoka.

There are still certain issues, such as the unnaturalness of translated phrases and cultural mismatch. However, the combination of machine translation and tourism English training provides an effective approach to prepare students to actual communication in international tourism environments.

Keywords

Machine Translation, tourism, tourism English training, airport

1 Introduction

In the past ten years, the number of people visiting Japan from overseas has grown quickly. Many come for sightseeing, study, or work, but communication is still not easy for them. When travelers go to smaller towns or countryside areas, they often find it hard to express themselves in English. At the same time, local shop staff and public workers sometimes struggle to answer properly in another language. Because of this, both sides may feel nervous or hesitate to communicate, even when they want to.

Recently, the wide use of machine translation (MT) tools has started to change this situation. Apps such

as Google Translate and DeepL are now familiar to most people, and many travelers use them every day on their phones. Modern versions include offline translation, voice recognition, and even camera-based text reading, which help users talk or understand signs without Wi-Fi. These tools make travelers feel more relaxed and independent, and they also make communication smoother for local people who want to help but cannot speak foreign languages well.

At the same time, MT is becoming more common in English as a Foreign Language (EFL) classes. Many students use translation tools to check vocabulary, understand reading materials, or review their writing. Some teachers worry that this may make learners too dependent on technology, but others believe MT can support learning when used carefully. It helps students notice grammar patterns, compare expressions between English and Japanese, and gain confidence to use English in real communication.

The present study looks at MT from two sides. First, it examines how travelers use translation tools in Japan and how this affects their behavior and confidence. Second, it explores how MT can be used in English learning to reduce barriers and support comprehension. By connecting tourism and education, this paper argues that MT can help create a more inclusive and practical communication environment in Japan—one where technology supports human understanding rather than replacing it.

2 Literature Review

Research on the use of machine translation (MT) in education has steadily increased over the past decade, alongside the growing presence of translation tools in everyday communication. Earlier studies tended to focus on technical aspects such as accuracy and system performance. More recent research, however, has shifted toward how MT is actually used by people and how it affects communication, learning behavior, and user confidence (O'Neill, 2019). Studies on second language learners suggest that students, particularly those with lower language proficiency, often hold positive attitudes toward MT. Rather than viewing it as a replacement for language learning, many learners see MT as a form of support that reduces anxiety and helps them participate in communication more actively. O'Neill (2019) reports that learners frequently use online translators to confirm meaning, check expressions, and gain reassurance when using a second language.

Research has also shown that MT use is not uniform among users. García and Peña (2011), for example, demonstrate that learners employ MT in different ways depending on their goals, ranging from quick comprehension to reflective comparison of expressions. Their findings suggest that MT can function as a learning aid when students are encouraged to notice differences between their own output and translated results, rather than simply accepting translations at face value.

In addition, the characteristics of MT tools themselves appear to influence how meaning is conveyed. Bowker and Ciro (2019) discuss how some MT systems prioritize structural or form-based equivalence, while others allow more flexibility in conveying contextual or pragmatic meaning. This distinction is important

because it can affect how cultural meaning is interpreted, particularly in real-world communication settings such as travel, service encounters, or intercultural interaction.

Taken together, existing research suggests that MT occupies a hybrid position between a communication tool and a learning resource. While it can lower linguistic barriers and increase user confidence, its educational value depends largely on how it is used and guided. Building on these insights, the present study focuses on Japanese learners and travelers, exploring how MT use—including differences between major tools—can support real-world communication while also revealing potential sources of cultural mismatch in interpretation.

3 Traveler Profile: Insights from Previous Studies

Previous studies on travelers' use of machine translation have suggested that younger users and those with lower language proficiency tend to show more positive attitudes toward translation tools (Bowker & Ciro, 2019), particularly as a means of reducing anxiety and facilitating communication during travel. Older travelers or those with higher education were generally more cautious. They trusted their own language skills and preferred to talk to people directly, often seeing MT as unnecessary or a bit awkward to use. This shows that age and education can strongly influence how people accept new technology.

Previous studies have also reported differences in language ability and communication behavior among travelers, particularly in situations where mobile technologies support travel experiences (Wang, Xiang, & Fesenmaier, 2016). Research on tourism information behavior also shows that travelers increasingly rely on digital tools and online resources when planning and experiencing travel (Xiang, Magnini, & Fesenmaier, 2015). Travelers with limited English skills or from minority-language backgrounds tended to rely on MT more often, using it as a kind of equalizer that helped them communicate smoothly in unfamiliar places. In contrast, those who already spoke English or another major language well used MT mainly as a backup when facing difficult phrases. Interestingly, even frequent MT users often said, "English is good enough." This attitude suggests that travelers care more about being understood and communicating effectively than about perfect grammar or accuracy. In the end, communication itself seems to matter most when people are abroad.

Research in language education has also discussed how translation technologies can support communication and language awareness when used appropriately (Garcia & Peña, 2011). The study also revealed differences in travel behavior. Those who depended on MT were more likely to visit nature spots or outdoor activity areas rather than large cities, perhaps because they were motivated by curiosity and a desire for new experiences. Meanwhile, older or more confident travelers preferred traditional destinations such as heritage sites or family-friendly spots. What stands out, however, is that using MT did not make travel less meaningful. Many MT users reported feeling more confident, curious, and proud of their experiences, showing that translation tools can actually support learning and intercultural growth.

The present study builds on these ideas. It aims to understand how both students and travelers in Japan use MT in daily life and what they think about it. While previous studies mainly focused on accuracy or technology, this research looks at how people use MT in real situations and how it affects their feelings and learning. Specifically, it asks: (1) Which MT tools are used most often? (2) For what purposes? and (3) How do users see MT in education? By collecting data from university students studying tourism and English, the goal was to connect real-world travel experiences with classroom learning. The results presented in the next section offer a clearer picture of how MT is used today and what it might mean for English education in Japan.

4 Machine Translation Infrastructure in Airport and Tourism Contexts

In recent years, airports around the world have started to adopt machine translation (MT) devices to improve multilingual communication. As international travel has become more diversified, it is no longer possible to rely only on English as a common language. Airports, which serve as the first point of contact between visitors and a country, have therefore introduced translation technologies to support both passengers and staff. These devices help travelers obtain information, ask questions, and solve problems quickly in situations where language barriers might otherwise cause confusion or stress.

Japan is one of the countries that has actively promoted such technological solutions. For example, All Nippon Airways (ANA) introduced POCKETALK® devices to more than 50 domestic airports. At Kansai International Airport, over 250 units of handheld translators are used to assist communication between ground staff and foreign travelers. Tokyo's Haneda Airport has also tested a transparent-screen translation system developed by TOPPAN Inc., which allows conversations to be displayed in multiple languages simultaneously. Similar developments can be seen across Asia. Shanghai Pudong International Airport, for instance, uses iFLYTEK translator machines that support more than 30 languages and can operate offline.

These examples show that MT is no longer limited to private smartphone use; it has become part of the infrastructure of international travel. Translation technologies in airports now play a central role in building inclusive, multilingual environments for travelers from all linguistic backgrounds. Table 1 summarizes the translation devices and systems deployed at major U.S. airports, along with the number of supported languages and their purposes. The wide range of supported languages ensures travelers can smoothly complete boarding procedures and navigate their way even if English is not their native language.

Table 1

Major Airport Implementations of Translation Devices in the U.S.

Airport	*Device / System*	*Languages Supported (Approx.)*	*Purpose*
JFK International Airport (Terminal 4, USA)	Google Assistant Interpreter Mode	~29 languages including Arabic, French, Japanese, Spanish (airport.nridigital.com)	Installed to reduce language barrier for international travelers; first airport terminal to deploy real-time translation tech.
Transportation Security Administration at Philadelphia International Airport (USA)	Hand-held translation devices (Pocketalk)	92+ languages (Pocketalk)	TSA testing large-scale deployment of hand-held translators for security checkpoint staff and passengers.
— (General Airport Market)	Real-time multilingual translation apps/devices	Market report shows translation-app deployment across airports globally (growthmarketreports.com)	Indicates the growing trend of airports adopting translation solutions for passenger assistance and service.

The following table summarizes major examples of airport translation-device implementations in Japan and China.

Table 2

Additional Airport Translation Device Implementations in Japan and China

Airport	*Device / System*	*Languages Supported (Approx.)*	*Purpose*
Japan (Domestic, 50 airports)	POCKETALK® hand-held interactive translator by All Nippon Airways (ANA)	Up to 74 languages (55 audio/text + 19 text-only) (Business Traveller)	Introduced by ANA to all 50 domestic airports in Japan to help ground staff communicate with international passengers.
Japan – Kansai International Airport (Kobe/Osaka)	POCKETALK® devices (250 units)	Handles ~70 languages (News On Japan)	Airport equipped staff with handheld translators to reduce communication problems before large tourist influx.

Table 2 (cont.)

Additional Airport Translation Device Implementations in Japan and China

Airport	*Device / System*	*Languages Supported (Approx.)*	*Purpose*
Japan – Haneda Airport, Tokyo	Transparent screen translation system (by TOPPAN INC.)	13 languages including English & Korean (Cybernews)	Pilot test at airport counters; uses a clear-screen system where translation appears on transparent display to support staff and travelers.
China – Shanghai Pudong International Airport	iFLYTEK Translator Machines at One-Stop Service Center for Foreigners	Recognizes 35 languages; covers over 200 countries/regions (iFLYTEK)	Deployed to assist international travelers in major hub airport in China; includes dual-screen design and offline translation.

Similar to the U.S., Asian countries like Japan and China also support a large number of languages, ensuring travelers can smoothly navigate their travel procedures even if they do not speak the local language. And then such developments also have implications for language education.

As students in tourism and hospitality programs prepare to work in airports or travel industries, exposure to these technologies in their studies can bridge the gap between classroom learning and real-world communication. Understanding how MT operates in actual service contexts may also help learners develop practical intercultural and linguistic skills that align with the demands of the global tourism industry. Such observations provide useful insights for language educators who seek to connect technological tools with practical communication training. They also highlight the importance of guiding students to use MT critically rather than relying on it unreflectively. Therefore, in the latter part of this paper, we will explore the implications for how machine translation can be utilized in classroom learning.

5 How Students Use Machine Translation and What They Learn from It

5.1 Frequency and Purpose of Use

The table below shows the main reasons why students at several Japanese universities use MT and their frequency of use.

Table 3

Reasons for Using Machine Translation

Main Reason	*Percentage*	*Description*
To understand written English texts	68%	Reading articles, academic texts, or signage
To communicate with foreigners	55%	Conversation or message translation
To support English learning	49%	Vocabulary building, comparing expressions
To reduce anxiety and uncertainty	46%	Boosting confidence when traveling or speaking
For convenience and speed	44%	Instant access to multilingual communication

Note. The second author's survey conducted among 328 university students from four Japanese universities — Kanagawa Dental University, Waseda University, Chiba University of Commerce, and Tokyo Seitoku University (2022).

The table reveals that Japanese university students utilize MT for a wide range of purposes, from building English vocabulary and expressions to reading comprehension and communication. They also highlighted *ease of use* and *psychological reassurance* as key benefits, aligning with the notion that MT can alleviate "language anxiety."

5.2 Educational Implications

Students reported that using machine translation (MT) helped them notice grammar points and vocabulary that they did not know before. By checking MT results, they could understand how words and sentence structures were used in real contexts. Many students also mentioned that MT made them more curious about language differences between English and Japanese.

Interestingly, learners who used MT more often seemed to develop a deeper awareness of how translation works. They were able to recognize small differences in meaning, tone, and expression, and sometimes tried to find better ways to express the same idea in English.

Overall, these results suggest that MT may play two roles for students: it can help them in practical situations, such as traveling or communication, and it can also be used as a reflective learning tool that encourages curiosity and deeper understanding of language.

6 Discussion

Machine translation, or MT, has started to change the way people deal with language, both in tourism and in education. It's not something new anymore, but the way people use it keeps changing. In the past, many users only turned to MT when they got stuck, just to check a word or phrase. Now, many of them use it more actively — to explore meaning, to test ideas, or simply to feel safe when communicating in another language.

For travelers, MT is, in a way, like a small travel partner. It helps them read signs, order food, and talk with people they meet. Some travelers said that having a translator in their phone made them feel less nervous,

which actually gave them the courage to explore more places. It doesn't solve every problem, of course, but it makes travel smoother and more human. In classrooms, MT plays a slightly different role. Students often use it to understand English sentences or to find words they don't know. But sometimes, while comparing their own sentences with MT results, they realize that their expression sounds too strong or not quite natural. That small discovery helps them think about how meaning changes with tone or culture.

Still, teachers need to be part of this process. Without guidance, students may copy everything from MT and lose the chance to think. So, teachers can ask questions like, "Why do you think this sounds better?" or "What happens if we change this word?" Such small moments make students more reflective and aware of language. MT, when used like this, becomes more than a shortcut — it becomes a kind of bridge between people and technology, between language learning and real communication.

7. Pedagogical Implications

7.1 Can Machine Translation Be Used in University English Classes?

These days, many travelers use SIM cards or free Wi-Fi to open machine translation apps like Google Translate or DeepL. People use them to talk with locals, read menus, or find their way when they get lost. It's kind of amazing how much easier travel has become because of that. And if MT already helps people in real life, then maybe it can help in English classes, too. This study thinks that, when teachers use it carefully, MT can make learning English more connected, more confident, and a bit more fun. There are a few simple ways MT can be used in class. The ideas below are not new, but they can make lessons feel closer to real communication.

1. Real-World Practice

Students can use MT during small role-plays that copy travel or service situations. They might act as hotel staff, or help a tourist in class. It's not perfect English, but that's okay—it's practice that feels real, not just from a textbook.

2. Support for Beginners

For students who still struggle, MT gives a little help. They can check meanings or instructions quickly, and that makes them less nervous. When they feel less afraid, they talk more, even in simple English. That small change matters a lot.

3. Building Critical Thinking

MT isn't always right. Sometimes it sounds strange, or it misses the feeling of the sentence. When students compare their own English with the MT version, they notice that difference. Then they start to ask questions—why this sounds better, why that sounds too formal—and that's real learning.

4. Preparation for Future Work

In tourism, hotels, or global business, translation apps are everywhere now. Students who know how to use them properly will be ready for that kind of job. Learning to use MT wisely at school means they can use it confidently later, not blindly.

8. Conclusion

In the end, machine translation can do more than just convert words. It opens space for real communication — during travel, in study, and in daily life. People often say it helps them feel less nervous when speaking, and somehow a little braver to try. In this way, MT doesn't only remove barriers; it gives people a small push to connect across different languages and cultures.

At the same time, MT now sits between language learning and technology. When teachers use it carefully, it becomes part of learning instead of a way to avoid it. Students who use MT slowly begin to notice patterns or small mistakes, and that makes them think about how language really works. They start asking questions: Why does this sound different? Why does that feel more natural? It's a quiet process, but it builds awareness. The value of MT may lie in this shift. It turns hesitation into curiosity, and dependence into confidence. Used in the right spirit, it doesn't weaken language learning—it adds to it. And maybe it reminds us, both teachers and learners, that studying English is not about perfection. It's about trying to understand and to be understood, one sentence at a time.

References

Bowker, L., & Ciro, J. B. (2019). *Machine translation and global research: Towards improved machine translation literacy in the scholarly community.* Routledge.

García, I., & Peña, M. I. (2011). Machine translation-assisted language learning: Writing for beginners. *Computer Assisted Language Learning*, 24(5), 471–487.

Japan National Tourism Organization. (2023). *Inbound tourism communication survey*. JNTO.

O'Neill, E. M. (2019). Online translator, dictionary, and search engine use among L2 students. *Computer Assisted Language Learning*, 32(1), 1–29.

Rumelhart, D. E. (1977). Toward an interactive model of reading. In S. Dornic (Ed.), *Attention and performance VI* (pp. 573–603). Academic Press.

Wang, D., Xiang, Z., & Fesenmaier, D. R. (2016). *Smartphones and the travel experience. Journal of Travel Research*, 55(8), 1010–1024.

Xiang, Z., Magnini, V. P., & Fesenmaier, D. R. (2015). Information technology and consumer behavior in travel and tourism: Insights from travel planning using the internet. *Journal of Travel Research, 54*(4), 511–527.

Integrating AI and Debate-based Role-Play: Enhancing EFL Speaking Through a Devil's Advocate Approach

Manami Sato

Kyoto University of Advanced Science

sato.manami@kuas.ac.jp

Abstract

This study explores an instructional approach integrating AI tools, debate, and role-play to enhance Japanese university students' speaking and argumentation skills. Although debate activities are known to improve fluency and confidence, many learners remain overly dependent on prepared scripts, limiting spontaneous and persuasive speech. To address this issue, generative AI tools such as ChatGPT were used to provide model speeches and rhetorical scaffolding, which students adapted for their own oral production.

In addition, a "Devil's Advocate" role was introduced, requiring students to argue positions contrary to their personal beliefs. This approach aimed to promote critical thinking and perspective-taking by engaging learners with opposing viewpoints. Micro-debates consisting of brief, timed exchanges were also incorporated to encourage quick thinking and interactive dialogue. Furthermore, the use of humor and exaggerated roles in role-play activities helped reduce psychological stress and supported more expressive communication.

Classroom observations and student reflections suggest that this integrated approach may have supported learners' perceived flexibility, confidence, and ability to construct nuanced arguments. The findings indicate a tendency for combining AI-assisted scaffolding with debate-based pedagogy to foster more adaptive communicative competence in EFL contexts, although these results should be interpreted as observations of classroom practice rather than objective performance gains.

Keywords

debate, devil's advocate strategy, generative AI, Japanese university context, speaking fluency

1 Introduction

In English as a Foreign Language (EFL) education, fostering learners' ability to engage in spontaneous, persuasive speaking remains a persistent challenge, especially in Japanese university contexts where classroom talk is often limited to prepared scripts or formulaic

responses. Traditional debate activities have been widely used to encourage critical thinking and oral fluency; however, students frequently rely on memorized discourse, which constrains their ability to interact adaptively in real communicative situations. This limitation highlights a gap in pedagogical approaches that simultaneously support fluency development and deep argumentation skills in real time.

To address this gap, this study integrates generative artificial intelligence (AI) tools with structured communicative activities, such as debates and role-play, to enhance learners' speaking competence. Generative AI, including large language models like ChatGPT, offers opportunities to scaffold students' rhetorical frameworks by providing model responses that learners can analyze and adapt. In classroom settings, AI can serve not only as a source of language input but also as a rehearsal partner, helping students internalize functional patterns of persuasive discourse.

A central innovation of this study is the adoption of the "Devil's Advocate" role within debate and role-play activities. Broadly defined, playing the Devil's Advocate involves intentionally adopting a position that counters the prevailing view in order to challenge assumptions and stimulate deeper analysis. The term originates from the historical role within canonization processes in the Catholic Church, where an appointed advocate argued against sainthood to uncover potential flaws, and in educational contexts it is used to promote critical inquiry by questioning dominant perspectives.

Educational research suggests that this strategy can counter groupthink and encourage students to consider alternative viewpoints, thereby enhancing critical thinking and communicative flexibility. In this study, students were introduced to the concept of Devil's Advocate through both explanation and media examples, prompting them to take on counterarguments in micro-debates and extended discussions. The combined use of AI scaffolding and role-play activities aimed to cultivate not only linguistic fluency and accuracy but also strategic competence and confidence in handling opposing positions.

2 Literature Review

2.1 Debate and Speaking Development in EFL Contexts

Debate has long been recognized as an effective pedagogical tool in EFL education for promoting learners' speaking fluency, confidence, and critical thinking. Through structured argumentation, learners are encouraged to articulate opinions, respond to counterarguments, and negotiate meaning in real time. Previous studies have shown that debate activities can enhance oral proficiency by increasing speaking opportunities and motivating learners to use language purposefully (Brown, 2014; Warschauer & Healey, 1998).

However, research has also noted limitations in debate-based instruction, particularly in contexts where learners tend to prioritize accuracy over communicative effectiveness. In Japanese EFL classrooms, students

often rely heavily on prepared scripts, which may improve grammatical accuracy but restrict eye contact, facial expressions, and spontaneous interaction. This tendency reflects a broader examination-oriented learning culture, where success is often associated with correctness rather than communicative impact. As a result, learners may struggle to transfer debate skills to authentic speaking situations that require flexibility and immediacy.

Moreover, group discussions frequently lead to opinion convergence, as learners avoid expressing minority views in order to maintain harmony. Such conformity can limit the depth of discussion and reduce opportunities for genuine critical engagement. These challenges suggest the need for instructional designs that not only promote debate participation but also explicitly encourage perspective-taking and spontaneous speech.

2.2 The Devil's Advocate as a Pedagogical Strategy

The Devil's Advocate strategy has been widely discussed in educational and organizational contexts as a means of preventing groupthink and stimulating critical discussion. By intentionally assigning a participant to argue against the dominant or majority view, this role encourages deeper examination of assumptions and consideration of alternative perspectives. In educational settings, adopting the Devil's Advocate role has been shown to promote analytical reasoning and more balanced deliberation.

From a sociocognitive perspective, arguing for a position that one does not personally endorse requires learners to detach from personal beliefs and focus on logical structure and evidence. This process can deepen conceptual understanding while also reducing affective barriers associated with personal disagreement. For language learners, such role distance may lower anxiety, as the speaker is not personally responsible for the opinion expressed, but rather performing an assigned role.

Despite its potential, the Devil's Advocate approach has received limited attention in second language speaking research, particularly in combination with debate and role-play. Existing studies tend to focus on critical thinking outcomes rather than linguistic development, leaving a gap in understanding how this strategy may influence fluency, accuracy, and complexity in L2 speech. This study addresses this gap by embedding the Devil's Advocate role within structured debate activities and examining its impact on learners' speaking performance and engagement.

2.3 AI-Supported Speaking Practice and Scaffolding

Recent advances in generative AI have introduced new possibilities for language learning, particularly in providing individualized and immediate support for learners. AI tools such as ChatGPT can generate model texts, suggest alternative expressions, and simulate conversational partners, offering learners opportunities for rehearsal and reflection beyond classroom time. Rather than replacing instructors, AI can function as a form of pedagogical scaffolding that supports learners' preparation and confidence.

Research on technology-mediated language learning suggests that such tools are most effective when integrated into task-based activities rather than used in isolation (Chong & Reinders, 2020). In speaking instruction, AI-generated models can help learners notice gaps between their intended meaning and actual production, thereby supporting metalinguistic awareness. However, concerns remain regarding overreliance on AI-generated language, particularly with respect to learners' spontaneous production and linguistic complexity.

From a cognitive perspective, Skehan's (1998, 2009) trade-off hypothesis provides a useful framework for examining these concerns. According to this model, learners' attentional resources are limited, and gains in fluency may come at the expense of accuracy or complexity, especially during spontaneous speech. The integration of AI support with debate and role-play therefore raises important questions regarding how different aspects of speaking ability develop and interact.

2.4 Research Gap and Focus of the Present Study

While previous research has examined debate instruction, critical thinking strategies, and AI-assisted language learning separately, few studies have explored their combined effects on EFL speaking development. In particular, there is a lack of empirical research investigating how AI-supported debate activities incorporating the Devil's Advocate role influence learners' fluency, accuracy, complexity, and willingness to engage with opposing viewpoints.

To address this gap, the present study examines a classroom-based intervention that integrates AI tools, debate, and role-play within a Japanese university EFL context. By analyzing learners' perceptions, classroom observations, and performance data, this study aims to contribute to applied linguistics research by offering pedagogical insights into how structured opposition and AI scaffolding can foster more adaptive and confident communicators.

3 Methodology

3.1 Participants and Context

The participants in this study were 28 first-year Japanese university students enrolled in a compulsory English course at a private university in Kyoto. All participants provided informed consent prior to data collection. The students were majoring in law and were part of a cross-disciplinary English curriculum consisting of two classes per week: Reading and Writing and Listening and Speaking.

Based on institutional placement data, the participants' English proficiency ranged from CEFR B1 to B2, with TOEIC scores approximately between 500 and 700. As is common in Japanese EFL contexts, the students had extensive experience with grammar-focused instruction and written examinations but limited opportunities for extended oral communication, particularly in argumentative or interactive speaking tasks.

3.2 Course Design and Instructional Framework

The course was conducted over a 15-week semester and integrated debate activities into a curriculum officially designated as a Reading and Writing course. This design intentionally expanded speaking opportunities within institutional constraints. Debate and discussion activities were introduced gradually, beginning with short, low-stakes tasks and progressing to more structured debates and group presentations.

Throughout the semester, students engaged in weekly debate sessions in pairs or small groups, adopting affirmative and negative positions on a variety of topics. Mid-term and final assessments included group debate presentations, allowing students to apply argumentative strategies in more extended speaking tasks. To reduce anxiety and encourage participation, humor and exaggerated roles were incorporated into role-play activities, particularly in micro-debate formats.

3.3 Debate Topics and Formats

Debate topics were selected to align with students' cognitive and linguistic levels and were categorized into three types:

(1) personal value topics (e.g., preferences in daily life),

(2) presumptive issues related to education and society, and

(3) broader policy-oriented issues.

This progression allowed students to move from familiar, low-risk topics to more abstract and socially complex issues. Several debate formats were employed during the course, including parliamentary debate for mid-term and final presentations, as well as modified micro-debates for regular classroom use. The micro-debate format consisted of short, timed exchanges designed to promote quick thinking, concise expression, and interactive engagement.

3.4 The Devil's Advocate Role

A central feature of the instructional design was the introduction of the Devil's Advocate role. Students assigned to this role were required to argue against the majority opinion or adopt a position opposite to their personal beliefs. The concept was introduced through explicit explanation and a short video illustrating the function and purpose of a Devil's Advocate in discussion settings. This introduction emphasized that the role was not intended to negate others' opinions, but to deepen discussion and explore alternative viewpoints.

In classroom practice, the Devil's Advocate role was implemented primarily within micro-debate activities and small-group discussions. In some cases, the role functioned as a third participant who challenged both sides; in others, it served as a judging or questioning role. Students were encouraged, but not required, to volunteer for this position. Over time, an increasing number of students actively chose to take on the role, suggesting growing acceptance and understanding of its value.

3.5 Use of AI Tools

Generative AI tools, specifically ChatGPT, were integrated into the course as a form of instructional scaffolding rather than as a substitute for learner production. Students used AI primarily during the preparation stage of speeches and group presentations to generate model arguments, organize ideas, and rehearse expressions. They were encouraged to adapt AI-generated content to their own language level and communicative style.

The use of AI was positioned as a support for confidence-building and rehearsal, particularly for prepared speeches. Students were also allowed to use AI for self-directed practice outside class, including checking clarity and experimenting with alternative phrasing. However, AI use was not permitted during in-class spontaneous speaking tasks, allowing for observation of learners' unaided oral performance.

3.6 Data Collection

Multiple sources of data were collected to examine the effects of the instructional intervention. These included post-course questionnaires, students' written reflections, classroom observations, and performance data from speaking tasks and presentations. The questionnaire consisted of both closed-ended items using Likert-scale responses and open-ended questions eliciting students' perceptions of debate activities, the Devil's Advocate role, and perceived skill development.

In addition, pre- and post-course Mini-TOEIC scores were collected to provide supplementary information regarding changes in general English proficiency. While TOEIC scores were not the primary measure of speaking development, they offered a reference point for interpreting classroom-based outcomes.

3.7 Analytical Framework

Data analysis focused on three primary dimensions of speaking performance: fluency, accuracy, and complexity. Learners' self-reports and reflective comments were analyzed thematically to identify perceived changes in confidence, engagement, and critical thinking. Observational data were used to triangulate these perceptions, particularly in relation to students' willingness to speak, vocal volume, use of gestures, and responsiveness during debates.

In the present study, these dimensions were primarily interpreted based on learner perceptions and classroom observations rather than direct linguistic measurements; therefore, the findings should be understood as indicative trends rather than objective performance gains.

Interpretation of the findings was informed by Skehan's (1998, 2009) trade-off hypothesis, which provided a framework for understanding potential tensions among fluency, accuracy, and complexity in learners' spoken production.

4 Findings

This section presents the findings of the study in relation to the three research questions. Data were drawn from post-course questionnaires, open-ended student comments, classroom observations, and performance-related records. The results are organized according to each research question.

4.1 RQ1: Effects of the Devil's Advocate Role on Critical Thinking and Perspective-Taking

RQ1: How does the Devil's Advocate role influence critical thinking and the exploration of different viewpoints?

The findings indicate that the introduction of the Devil's Advocate role had a positive impact on students' engagement with multiple perspectives and the depth of classroom discussions. Questionnaire responses showed high levels of interest in this role, with the majority of students rating it as "very interesting" or "interesting." Notably, no students reported negative perceptions of the activity.

Classroom observations revealed that discussions became more dynamic and balanced when a Devil's Advocate was present. Rather than quickly converging on a shared opinion, students were prompted to reconsider assumptions and provide additional justification for their arguments. This shift was particularly evident in micro-debate activities, where the Devil's Advocate challenged both affirmative and negative positions, resulting in extended interaction and clarification requests.

Qualitative data from student reflections further illustrate this effect. Several students commented that discussions tended to stagnate when all participants shared the same opinion, whereas the presence of a Devil's Advocate encouraged deeper and more meaningful exchanges. Others noted that the role helped them recognize the value of opposing views and understand that disagreement could enhance, rather than hinder, productive discussion.

In addition, an increase in voluntary participation was observed over the course of the semester. Initially, students were hesitant to adopt minority positions; however, as the purpose of the role became clearer, more students actively volunteered to act as the Devil's Advocate. This suggests that the role reduced conformity pressure and created a classroom environment in which expressing alternative viewpoints was normalized.

4.2 RQ2: The Role of AI Tools in Supporting Prepared and Spontaneous Speech

RQ2: How can AI tools support students' prepared and spontaneous speeches?

The findings suggest that AI tools were particularly effective in supporting prepared speaking tasks, such as group presentations and structured debates. Students reported that AI-generated model texts helped them organize arguments, identify appropriate expressions, and reduce anxiety related to grammatical accuracy. As a result, students appeared more confident during presentations, demonstrating improved vocal volume, clearer articulation, and greater use of eye contact and gestures compared to previous cohorts.

Student reflections indicate that AI use functioned as a rehearsal aid rather than a replacement for thinking. Many students described using ChatGPT to confirm whether their intended meaning was conveyed clearly and to explore alternative phrasing at an appropriate proficiency level. This process appeared to enhance noticing and awareness of rhetorical structure, contributing to smoother delivery in prepared tasks.

In contrast, the impact of AI on spontaneous speech was indirect. While AI was not used during in-class spontaneous speaking activities, students reported that prior rehearsal with AI reduced psychological stress, allowing them to focus more on interaction during debates. Classroom observations support this claim, as students demonstrated increased willingness to speak and respond under time pressure, particularly in micro-debate formats.

Overall, the findings suggest that AI tools played a supportive role in building confidence and preparedness, thereby facilitating more active participation in both prepared and spontaneous speaking contexts.

4.3 RQ3: Effects on Fluency, Accuracy, and Complexity

RQ3: How does the integration of AI, debate, and role-play affect fluency, accuracy, and complexity in learners' speaking?

The results may indicate perceived improvements in fluency, accuracy, and complexity, based on student feedback and classroom observations rather than objective linguistic analysis.

Specifically, regarding fluency, both self-reported data and classroom observations indicate noticeable improvement. A majority of students selected speaking-related skills as those that improved the most during the semester. Observational notes highlighted increased speech length, reduced hesitation, louder voices, and more expressive delivery, particularly during interactive speaking tasks.

In terms of accuracy, students perceived improvement primarily in prepared speech contexts. AI-supported drafts were largely error-free, allowing students to focus less on form monitoring and more on delivery. However, this does not necessarily indicate acquisition-level improvement in grammatical accuracy, but rather a reduction in cognitive load during performance.

With respect to complexity, the findings suggest a tendency toward simplification in spontaneous speaking tasks. Students often relied on shorter sentences and more basic structures when responding under time constraints. This pattern aligns with Skehan's (1998, 2009) trade-off hypothesis, which posits that gains in fluency may come at the expense of complexity when attentional resources are limited.

Supplementary data from Mini-TOEIC scores showed a slight decrease from pre-test to post-test. While this result may appear contradictory, it reflects the difference between standardized test performance and classroom-based communicative development. The findings suggest that increased emphasis on speaking and

interaction did not immediately translate into improved test scores, particularly for freshmen adjusting to university-level learning after intensive entrance examination preparation. This decline may also reflect test-taking fatigue or the shift in instructional focus from test-oriented learning to communicative practice.

4.4 Summary of Findings

Overall, the findings indicate that the integrated use of AI tools, debate, and role-play—particularly the Devil's Advocate role—contributed to enhanced learner engagement, increased fluency, and greater openness to alternative viewpoints. At the same time, the results highlight trade-offs among different dimensions of speaking ability and underscore the importance of interpreting proficiency development within both cognitive and contextual frameworks.

5 Discussion

This study examined the effects of integrating AI tools, debate, and role-play—particularly through the use of the Devil's Advocate role—on Japanese university students' EFL speaking development. The findings provide several insights into how structured opposition and AI-supported scaffolding can influence fluency, accuracy, complexity, and learner engagement in classroom-based communicative activities.

5.1 Devil's Advocate and the Development of Critical Engagement

One of the most salient findings concerns the impact of the Devil's Advocate role on critical thinking and classroom interaction. The results suggest that assigning students to argue from a minority or opposing position effectively reduced conformity pressure and encouraged more balanced and sustained discussion. In contexts such as Japanese EFL classrooms, where harmony and agreement are often prioritized, this structured role appears to legitimize disagreement as a productive and necessary component of communication.

From a pedagogical perspective, the Devil's Advocate role functioned as a form of cognitive distancing. Because students were not required to defend their personal beliefs, they could focus on constructing logical arguments and responding to counterpoints without the affective burden of personal confrontation. This role-based distancing likely contributed to increased participation and willingness to speak, particularly among students who were initially reluctant to express dissenting opinions.

These findings extend previous research on debate and critical thinking by demonstrating that explicitly assigning oppositional roles can deepen deliberation and increase interaction in L2 speaking tasks. While prior studies have emphasized debate as a means of fluency practice, this study highlights the importance of role design in shaping the quality and depth of communicative engagement.

5.2 AI as Scaffolding Rather Than Substitution

The findings provide insights into the pedagogical role of generative AI in speaking instruction. Rather than functioning as a substitute for learner production, AI tools in this study appeared to serve primarily as scaffolding for preparation and rehearsal. Based on student reflections, providing level-appropriate and structurally coherent model texts seemed to reduce learners' anxiety related to grammatical accuracy and organization, thereby potentially allowing greater attentional resources to be allocated to delivery and interaction.

This function aligns with previous research on technology-mediated task-based language teaching, which emphasizes the importance of integrating technological tools into meaningful communicative tasks (Chong & Reinders, 2020). The present findings suggest that AI may be most beneficial when its use is centered on preparatory stages, supporting confidence-building without diminishing opportunities for spontaneous language use.

Importantly, the study also highlights potential risks associated with AI use. While AI-generated drafts were linguistically accurate, they did not necessarily translate into increased complexity in spontaneous speech within this classroom context. This underscores the need for careful instructional design that encourages learners to move beyond reliance on pre-constructed language and engage in real-time language processing.

5.3 Interpreting Fluency, Accuracy, and Complexity Through the Trade-Off Framework

The observed patterns of speaking development can be meaningfully interpreted through Skehan's (1998, 2009) trade-off hypothesis. The marked improvement in fluency, alongside stable or reduced complexity in spontaneous speech, suggests that learners prioritized communicative effectiveness under time pressure. This prioritization is consistent with the cognitive demands of debate and micro-debate activities, which require rapid response and interaction.

The apparent discrepancy between classroom-based speaking gains and slight declines in standardized test scores further highlights the distinction between communicative competence and test-oriented proficiency. For first-year university students transitioning from entrance examination-focused learning, increased emphasis on interaction and expression may temporarily disrupt test performance while fostering longer-term communicative development.

Rather than viewing these trade-offs as deficiencies, they should be understood as indicative of learners' shifting attentional priorities. As fluency and confidence increase, opportunities may emerge for subsequent development in complexity and accuracy, particularly with sustained practice and targeted feedback.

5.4 Pedagogical Implications for EFL Instruction

The findings of this study offer several implications for EFL pedagogy. First, incorporating structured oppositional roles such as the Devil's Advocate can effectively promote critical engagement and reduce passive agreement in classroom discussions. Second, AI tools can play a valuable role in supporting preparation and rehearsal, particularly for learners who lack confidence in their linguistic accuracy.

Finally, the integration of debate, role-play, and AI highlights the importance of designing speaking tasks that balance support and challenge. By combining scaffolding with time-pressured interaction, instructors can create learning environments that foster fluency, confidence, and adaptability—key components of communicative competence in academic and professional contexts.

6 Limitations and Conclusion

6.1 Limitations

Several limitations of this study should be acknowledged. First, the research was conducted in a single institutional context with a relatively small sample size of 28 first-year university students. As a result, the findings may not be generalizable to other EFL contexts, proficiency levels, or educational settings. Future research involving multiple institutions or longitudinal designs would provide a more comprehensive understanding of the effects observed in this study.

Second, much of the data relied on self-reported perceptions and classroom observations. While these data sources offer valuable insights into learners' experiences and engagement, they do not provide objective measures of speaking development. More fine-grained analyses, such as acoustic analysis of speech or detailed discourse analysis, would allow for a more precise examination of changes in fluency, accuracy, and complexity.

Third, although AI tools were integrated thoughtfully as preparatory scaffolding, the study did not systematically compare AI-assisted and non-AI-assisted conditions. Future studies could employ experimental or quasi-experimental designs to more clearly isolate the effects of AI use on different dimensions of speaking performance.

Finally, the observed discrepancy between classroom-based communicative gains and standardized test results highlights the need for careful interpretation. While TOEIC scores provided supplementary information, they may not adequately capture changes in interactive speaking ability, particularly in short-term classroom interventions.

6.2 Conclusion

This study investigated the pedagogical effects of integrating AI tools, debate, and role-play—specifically through the use of the Devil's Advocate role—on Japanese university students' EFL speaking development. The findings suggest that this integrated approach enhanced learner engagement, promoted fluency, and encouraged critical engagement with opposing viewpoints, even within the constraints of a curriculum officially focused on reading and writing.

By assigning students to structured oppositional roles, the Devil's Advocate approach helped reduce conformity pressure and fostered deeper discussion, while AI tools supported preparation and confidence without replacing learner production. Interpreted through a cognitive trade-off framework, the results highlight how gains in fluency may coexist with stable or reduced complexity during spontaneous speech, particularly in time-pressured tasks.

Overall, the study contributes to applied linguistics research by demonstrating how AI-supported debate pedagogy can be used to develop adaptive communicative competence in EFL classrooms. Rather than emphasizing linguistic perfection, this approach prioritizes flexibility, confidence, and critical engagement—skills essential for meaningful communication in academic and professional contexts.

References

Brown, H. D. (2014). *Principles of language learning and teaching* (6th ed.). Pearson Education.

Chong, S. W., & Reinders, H. (2020). Technology-mediated task-based language teaching: A qualitative research synthesis. *Language Learning & Technology, 24*(3), 70–86.

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

Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics, 30*(4), 510–532.

Warschauer, M., & Healey, D. (1998). Computers and language learning: An overview. *Language Teaching, 31*(2)*,* 57–71.

Effects and Challenges Following COIL Education Implementation—Analysis Based on INTER-UNIVERSITY EXCHANGE PROJECT in 2018 (COIL project)

Ikuko Motohisa[1] & Emiko Koyama[2] & Yukiko Ideno[3]

[1]Chiba University, [2]Shibaura Institute of Technology, and [3]Tokiwa University

imotohisa@faculty.gs.chiba-u.jp, Koyama.emiko.h6@sic.shibaura-it.ac.jp, y-ideno@tokiwa.ac.jp

Abstract

Collaborative Online International Learning (COIL) promotes internationalization by enabling online collaboration among students from diverse cultural and linguistic backgrounds. While the approach originated in the State University of New York (SUNY) in 2004, it was introduced in Japan only recently. This study analyzes the Ministry of Education, Culture, Sports, Science and Technology (MEXT) reports on the “2018 MEXT COIL Project” within the Inter-University Exchange Project (IUEP) to investigate factors supporting sustainable COIL implementation in Japanese universities. By comparing the project outcomes of 13 higher education institutions, the study identified the strategies for successful implementation and the common challenges. Especially, universities that improved their evaluation from “A” to “S” tended to institutionalize outcomes, expand partnerships, and develop scalable models. The research findings suggest that a stable network of partner universities and a continuous search for new development opportunities are key to sustaining COIL initiatives in Japan.

Keywords

COIL, internationalization, partner institutions, the strategies of sustainable implementation, and challenges

1 Introduction

The increasing demand for internationalization in higher education began in the mid-1990s. While Japan was once globally competitive, it has since fallen to 38th place in 2024 (MEXT, 2025, p 50), partly due to a lag in digital competitiveness (MEXT, 2025, p 53). This decline necessitates "internal internationalization" efforts, which involve welcoming international students into educational programs.

Collaborative Online International Learning (COIL) has emerged as a global approach to fostering "internationalization at home". COIL involves international online interactive classes and new study abroad formats. In 2018, Japan adopted the COIL Fiscal Year 2018: Support for Establishing Inter-University Exchange Programs (IUEP) with the United States (U.S.) and other countries using the COIL model through IUEP. Although the benefits of COIL are becoming apparent (Ikeda, 2014; Inoue et al., 2022), its implementation methods and challenges remain unclear. Research on COIL in Japan is fragmented, with planning often left to individual universities.

This study aims to address these issues by analyzing reports from 13 universities that participated in the 2018 MEXT COIL-type projects. The analysis seeks to uncover implementation planning challenges and identify key success factors.

2 Literature Review

2.1 The Origins of COIL Worldwide

COIL is an educational approach where students from different countries and cultures collaborate online to address shared challenges. Originating in the U.S. about 20 years ago and introduced at the SUNY COIL center in 2006, COIL was developed for students facing difficulties with traditional study abroad due to economic disparities or limited local interaction opportunities. This program cultivates the essential skills for a diverse society: cross-cultural competence, communication skills that transcend language barriers, problem-solving skills, self-directed learning, and information and communication technology (ICT). COIL expanded from the U.S. to Europe and other regions due to the recognized importance of these skills in a globalized era and the growing demand for more students to gain international experience. Key factors driving this expansion include recognizing that these skills are essential in an era of globalization and the growing demand in every country for more students to study abroad and gain overseas experience.

The focus is on the COIL (Collaborative Online International Learning) concept for developing an international curriculum (Rubin & Guth, 2016). O'Dowd (2021) examined the definition of "meaningful interaction" and practical approaches based on social constructivism. The study analyzes the educational benefits and challenges of online interaction when in-person instruction is difficult and explores the potential of international education. Around 2000, amid advancing globalization, Japan's lack of international competitiveness became increasingly apparent, prompting many companies to urgently cultivate talent capable of succeeding globally. In response, MEXT launched the IUEP of *Japan Society for the Promotion of Science* (JSPS) in 2011, which led to the creation of COIL-type projects in 2018 (JSPS, 2018a; MEXT, 2018a).

2.2 The Start and Background of COIL in Japan

In Japan, COIL began at Kansai University in 2014 (Ikeda, 2020), and it has since expanded and become firmly established in higher education through initiatives such as the 2018 COIL project by IUEP and the JPL-COIL Association. Despite recent growth in research on COIL in Japan, studies remain limited, especially concerning its effectiveness (Nagata, 2020). Generalization is difficult due to varied methodologies, participant characteristics, and partnerships (Inoue, M. et al., 2022). Challenges persist, and comprehensive countermeasures are still underway. COIL is crucial for fostering cross-cultural exchange and international experience for both students and teachers, while allowing students to remain enrolled at their home institutions.

Since the early 2000s, Japan's decline in global competitiveness has driven a push toward internationalization, making the development of globally minded talent an urgent priority. Companies now seek graduates with strong communication skills and international experience (MEXT, 2025, p 55). Universities are seen as the final opportunity for students to acquire international experience before entering the workforce. The COVID-19 pandemic further accelerated the shift towards online learning and COIL models. This study aims to identify factors for sustainable COIL implementation in Japanese higher education. It addresses three research questions:

RQ1: What institutional characteristics are linked to successful COIL implementation in the 2018 MEXT COIL project?

RQ2: How do partner university networks relate to MEXT project evaluation outcomes?

RQ3: What challenges and sustainability factors are evident from implementation reports of participating universities between the 2018 project and the 2023 project ongoing?

To answer these questions, the study will conduct a comparative analysis of reports from the 13 higher education institutions that participated in the 2018 MEXT project (COIL-type Project), and track and analyze several universities that continued COIL projects after the 2023 MEXT project.

3 Methods

3.1 Research materials

This study employed a qualitative comparative content analysis of the mid-term and final reports submitted by 13 universities that participated in the 2018 MEXT COIL project. In addition, the reports of three universities that continued to participate in the 2023 MEXT COIL project are used. Each project received five-year funding from the Japan Society for the Promotion of Science (JSPS). The documents used for analysis can be downloaded from the JSPS website.

3.2 Procedure

The current study focuses on the following points:

(1) Rating changes between mid-term and final term for each university participating in the MEXT 2018 COIL-type project

(2) The number of international partner universities in the MEXT 2018 COIL project and its relation to mid-term and final evaluations

(3) Characteristics of the universities that received the highest S rating in the final report for MEXT COIL 2018

(4) The number and geographic region of partner universities between the MEXT COIL 2018 and MEXT COIL 2023 projects

First, to examine how ratings changed between mid-term and final for each university (or university groups), data from the 2018 mid-term and final reports were compared across three groups based on MEXT ratings: institutions that improved their ratings, those that maintained their ratings, and those that showed decline or limited advancement. This comparison enabled the identification of recurring patterns associated with successful implementation and longer-term sustainability. Second, to assess the impact of COIL implementation on the participating institutions, we examined the number of international partner universities in the 2018 COIL project. Content analysis was conducted to examine how the number of partner universities may relate to changes in MEXT ratings. Third, to examine the reasons for success, detailed content analysis of the mid-term and final reports of the highest-ranking universities was performed and categorized into (1) issues and challenges, (2) improvements and achievements, and (3) factors that may have affected the evaluation. Finally, three of the original 13 universities that continued their participation in both the 2018 and 2023 COIL projects were used to assess the long-term impact of COIL implementation. Changes across geographic regions between the 2018 and 2023 COIL projects were examined to assess the global impact of long-term COIL implementation.

Data coding was completed by one author and checked by another. To enhance methodological transparency, analytic memos were kept throughout the coding and categorization process. In addition, the interpretation of categories was cross-checked against previous COIL-related studies and relevant theoretical perspectives on international collaborative learning. This procedure was intended to improve the credibility of the analysis and to reduce overly impressionistic interpretation.

3.3 Data analysis

All descriptive analyses and graphical representations were completed using R Studio (version 2023.09.1+494). The coded categories included the name of the participating university (or groups of universities), Type of COIL project (Type A: Projects that promote international exchange, Type B: Project that promotes international exchange and that establishes an information platform for Japanese universities), the names of partnering universities, geographic regions of the partner universities, the number of participating students

(Japanese or international), and the mid-term/final evaluations. Comparative content analysis was conducted to identify similarities and differences between the midterm and final evaluations of the 2018 MEXT projects and across the ongoing 2023 MEXT project. The analysis focused on institutional strategies, the number of partner institutions, and program implementation characteristics described in the reports. Through iterative comparison of these elements, patterns associated with successful implementation and sustainability were identified.

4 Results

4.1 Changes from Mid-Term to Final Evaluations in the 2018 MEXT COIL Project

Figure 1 shows the changes in evaluation ratings for participating institutions in the 2018 COIL-type projects. The participating institutions and their university type (i.e., national, public, private) are shown in Table 1. The ratings ranged from A- to S. Changes from the mid-term to the final evaluation could be broadly categorized into three patterns. First, the universities that *improved* their MEXT ratings from A to S were Nanzan University, Tokyo University of the Arts, and the collaborative partnership between Tokyo University of Foreign Studies (TUFS) and International Christian University (ICU). The University of Tokyo was the only university to improve its rating from A- to A. Second, the universities that *maintained* the same A rating in both mid-term and final evaluations were the tri-university collaboration group consisting of Sophia University, Ochanomizu

Figure 1

Changes in Mid-term and Final Evaluations in the 2018 MEXT COIL Project

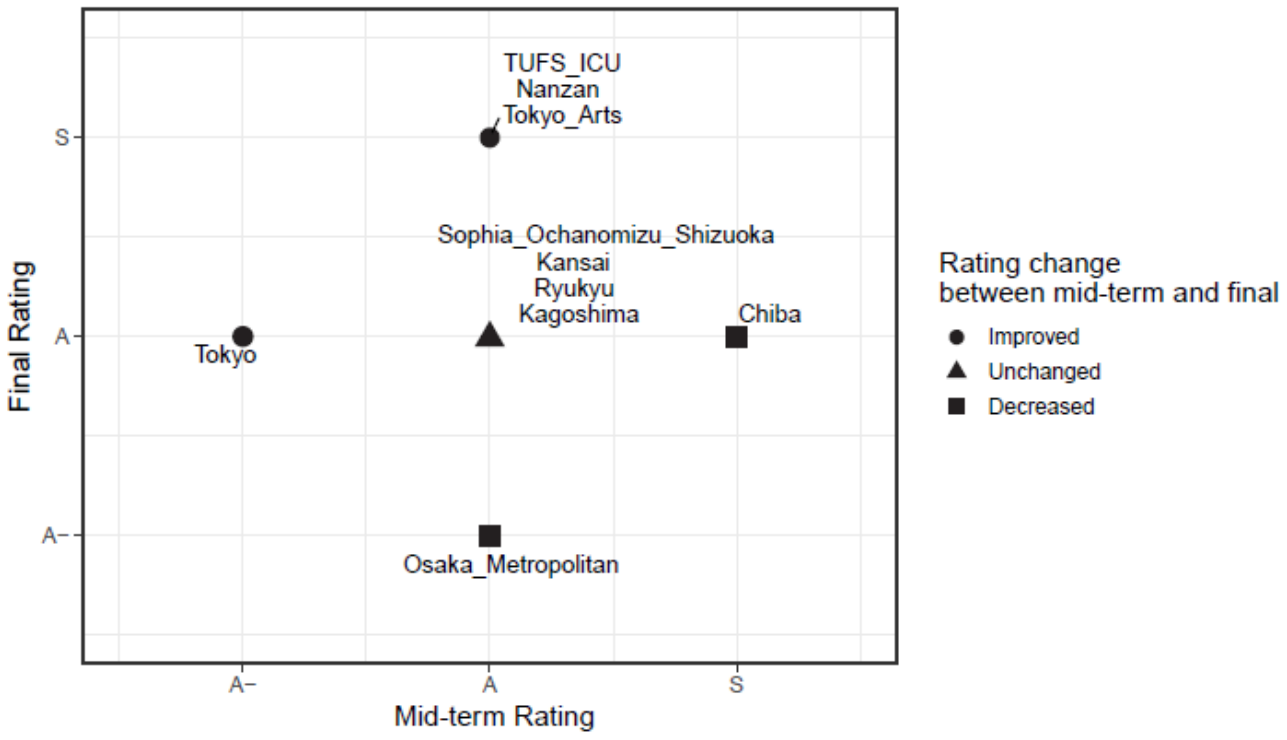


Table 1

Participating institutions in the 2018 COIL Project

University Type	University Name
National	Chiba University
National	Kagoshima University
National	Ochanomizu University
National	The University of Tokyo
National	Tokyo University of Foreign Studies (TUFS)
National	Tokyo University of the Arts
National	University of the Ryukyus
Public	Osaka Metropolitan University
Public	University of Shizuoka
Private	International Christian University (ICU)
Private	Kansai University
Private	Nanzan University
Private	Sophia University

University, and University of Shizuoka, as well as Kansai University, University of the Ryukyus, and Kagoshima University. Third, the universities whose final evaluations *decreased* from the mid-term evaluations were Chiba University and Osaka Metropolitan University. These three patterns were named “Increased”, “Unchanged”, and “Decreased” and used for the following analysis.

4.2 The Number of Partner Universities by Evaluation Change Group in the 2018 MEXT COIL Project

Figure 2 shows the number of partner institutions per evaluation change group (i.e., "Improved," "Unchanged," or "Decreased") between the midterm and final evaluations in the 2018 MEXT project reports, as described above. Both the "Improved" and "Unchanged" groups had a high number of partner institutions, averaging over nine institutions. Tokyo University of the Arts was an exception, with only one partner. Despite its limited international collaboration, a content analysis revealed that Tokyo University of the Arts conducted projects with industry partners and others. The university carefully provided students with feedback, making it a good example of prioritizing quality.

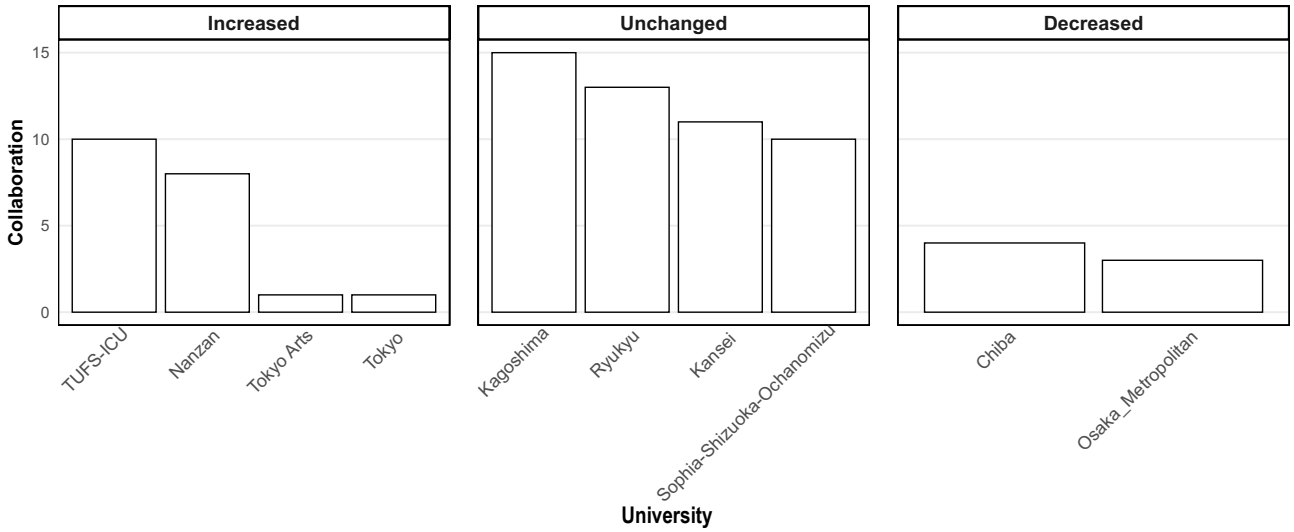
Conversely, the group whose evaluation declined had few partnerships. For instance, Chiba University had a large number of participating students and received an “S” rating in the midterm report. However, the results of the comparative content analysis revealed that failing to establish partnerships with planned domestic national universities and local institutions, as well as having a small number of partner institutions, may have

contributed to the decline in their final evaluation. Similarly, Osaka City University partnered with only two educational institutions in the United States. The content analysis showed that the university initially undertook ambitious projects that went beyond the scope of COIL; however, these initiatives did not result in COIL-related collaborations. The small number of partner institutions was likely due to insufficient substantive organizational activities, making it difficult to address the situation later.

The University of Tokyo likely faced a similar situation, despite its ratings improving from A- to A between the midterm and final reports. Initially, collaboration with MIT was planned, but organizational restructuring necessitated reaching out to other institutions separately.

Figure 2

The Number of Partner Universities by Evaluation Change Group in the 2018 MEXT COIL Project



4.3 Characteristics of Universities with S Ratings in the Final Evaluation in the 2018 MEXT COIL Project

Table 2 summarizes the characteristics of institutions that received the highest S ratings in the final evaluation for the 2018 MEXT COIL project. All these institutions received an A rating in the mid-term evaluation, but improved to an S in the final evaluation. The following points can be identified as characteristics of this group from the table. The commonality among these universities is that their mid-term evaluations identified "institutionalization, collaboration, and sustainability as challenges." On the other hand, their final evaluations are attributed to overcoming challenges such as "advancing institutional frameworks, expanding evaluations, and establishing sustainability."

Another characteristic is that each university possesses distinct strengths. For example, Nanzan University's strengths include institutional development, foundation-building, and maintaining exchanges. Tokyo University of the Arts' strengths lie in expanding its educational programs, increasing its social outreach, and becoming an international model. Finally, the strengths of Tokyo University of Foreign Studies and

International Christian University were identified as quantitative expansion of international education, model development, and expansion into new regions.

Table 2

S-Rated Universities: Evaluation Step-Up Comparison (2018 MEXT COIL Project)

University	Issues Identified in the Mid-term Evaluation (A)	Improvements and Achievements in the Final Evaluation (S)	Factors Leading to Evaluation(S)
Nanzan	- Need for fair assessment standards through performance analysis - Increasing diversity among international students - Maintaining relationships with partner universities - Managing reintegration of returning students - Strengthening overall organizational financial stability	- Standardization and credit recognition of COIL-based courses - Operation of three types of COIL programs - Promotion of collaborative tasks and information sharing among students - Establishment of an on-campus infrastructure for ongoing COIL activities - Increase in the number of exchange students to/from partner institutions	- Development of systems and organizational foundations -Maintenance of exchange activities
Tokyo University of the Arts	- Degree program and credit recognition not yet fully established - Strengthening collaboration with domestic universities - Challenges in securing faculty resources	- Implementation of short-, mid-, and long-term intensive programs - Establishment of a learning cycle with remote collaboration - Acceleration of students' inbound/outbound mobility - Demonstration of results through portfolio creation - Creation of an international education model	- Advancement of educational programs - Social outreach and establishment of a universal model

Table 2 (cont.)

S-Rated Universities: Evaluation Step-Up Comparison (2018 MEXT COIL Project)

TUFS and ICU	- COIL participation numbers below target - Unclear connection between courses and COIL - Need to strengthen pre- and post-program follow-up - Insufficient system for integrating COIL with regular courses - Lack of overall program cohesion	- Campus-wide course alignment with COIL - Nationwide expansion of COIL partnerships - Development of a comprehensive COIL curriculum - Promotion of regional integration involving Africa, Asia, and the Pacific	- Expansion of international education - Establishment of a model and creation of new development possibilities

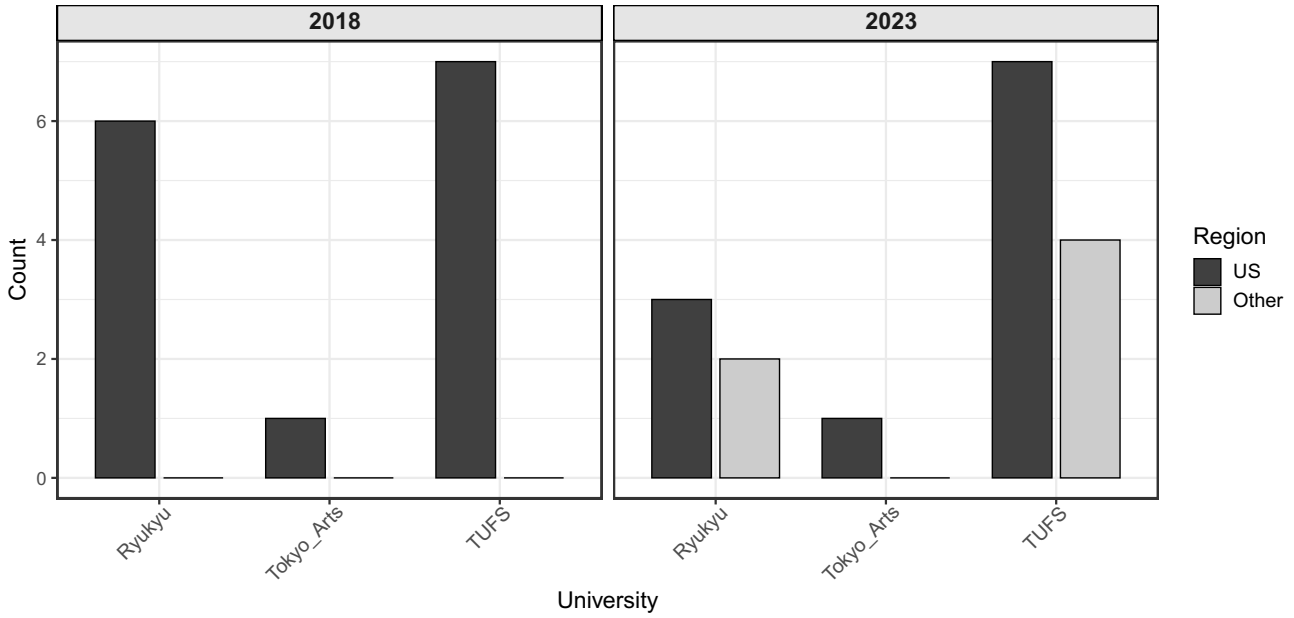
4.4 Changes in University Partnerships Between the 2018 and 2023 MEXT COIL Projects

Figure 3 compares the number of partner universities and their geographic distribution among the three universities that have continued to implement COIL in both the concluded 2018 project and the ongoing 2023 Type A projects (JSPS, 2026a; 2026b; 2026c). These data were analyzed to evaluate the long-term impact of COIL implementation in Japan.

Throughout the 2023 academic year, the Tokyo University of the Arts maintained its number of partner institutions. This is believed to be due to the meticulous implementation of the COIL project in 2018, which led to high-quality, rather than high-volume, activities. This indicates that the university continues to maintain strong partnerships with its partner institutions. The Tokyo University of Foreign Studies kept the same number of partner institutions in the United States while adding new ones in other regions, including Asia, such as the Philippines and South Korea. Conversely, the University of the Ryukyus halved the number of partner institutions in the United States, but added new ones from other regions. Overall, there is a clear trend toward an increase in the number of partner universities outside of North America, and these universities likely have unique reasons for doing so, such as establishing cooperative frameworks with island regions in line with their future goals.

Figure 3.

Geographic Regions (US / Other regions) of Partnership Universities in the 2018 and 2023 MEXT COIL Projects



5 Discussion and Conclusion

In this study, we classified universities in the COIL project into three groups based on changes in the mid-term and final evaluations by MEXT. The results of the analysis for each RQ are discussed below.

Regarding RQ1, many universities received an “A” rating in either the midterm or final evaluation. However, only three universities received an “S” rating in the final evaluation. An analysis of the three universities' common keywords revealed that "establishment of a model" was included. This suggests that an "S" rating indicates not only the implementation of COIL at each university but also the ability to generalize such learning activities and institutionalize them at other universities. Since these are not subjective evaluations, but rather terms common to all three universities, they can be considered qualities required at the "S" level. This point likely served as a clue to the keywords that raised the evaluation, rather than the implementation reports from individual institutions in previous studies.

Regarding RQ2, both the AA “unchanged” and AS “improved” groups had an average of 9 collaborative partners. This suggests that a certain number of partnerships is necessary for COIL evaluation and outcomes. Interestingly, the group that maintained an A-level rating in the final evaluation had more partner universities than the group that improved to an S-level rating. While a larger number of partnerships is generally advantageous, an examination of the relationship between groups whose ratings improved and those whose ratings declined suggests that a certain number of partnerships is required for stable operations—a condition met by the group that maintained an A-level rating. On the five-year project, having too few partner universities was considered risky because it made responding to unforeseen circumstances difficult. These findings are consistent with previous studies on COIL sustainability, which emphasize the importance of institutional

commitment and stable international partnerships (O'Dowd, 2021). Establishing long-term collaboration frameworks enables universities to sustain COIL activities beyond the completion of government-funded initiatives.

Several challenges and factors affecting stability were identified during the implementation of the COIL project regarding RQ3. For instance, in groups where mid-term and final evaluations decreased, the results of short-term collaborative learning did not carry over to long-term evaluations. Although the number of partner universities was around 9, Tokyo University of the Arts' experience clearly demonstrates that achieving high evaluations in the AS model ultimately depends on the quality of collaborative learning and model development, not just on quantity. In any case, it was evident that each university's initiatives in the AS model leveraged the characteristics of its region, faculty, and partner universities. Their modeling and unique approaches had a ripple effect on other institutions.

The analysis also indicates that universities engaged in these projects tended to collaborate with institutions with which they had already established relationships. Although rapidly expanding the number of partner institutions may be difficult, it is feasible to begin by implementing small-scale collaborative learning activities with existing partners and gradually expand the network by adding another area for each faculty member at their own university. These insights emerged from the analysis of institutional reports and related materials, particularly in comparison between universities such as the University of Tokyo, which began with an A− evaluation, and Chiba University, whose evaluation declined from S to A, and institutions that maintained A or improved to S. Further disclosure of comprehensive project data would make it possible to understand broader trends in global higher education better and to identify gaps and future directions in collaborative international learning during that period.

Acknowledgements

The original study was presented at the 29th PAAL International Conference in 2025. This paper presents new research that incorporates additional analyses of reports from the 2018 and 2023 MEXT COIL projects. We would like to express our sincere gratitude to the two anonymous reviewers for their valuable comments.

References

Ikeda, K. (2020). ICT o katsuyo shi kaigai no gakusei to okonau kokusairenkeigata no kyodo gakusyu "COIL" no kyoikukoka to kadai (in Japanese) [Educational Effectiveness and Challenges of COIL, an International Collaborative Learning Program Using ICT with Overseas Students], *JUCE Journal, 2*, 20-25

Inoue, M. Ohe, N. & Mano, K. (2022). Gurobaru PBL to saiba hwijicaru kuukan eno kakudai (in Japanese) [Global Project-Based Learning and its Extension to Cyber-Physical Space-Current and Future of

Cooperative Online International Learning (COIL) in Science and Engineering-] Journal of JSEE, 70-3,2022

Japan Society for the Promotion of Science. (2018a). COIL *gata kyoiku wo katsuyoshita beikokutou tono daigakukan koryukeisei shien* [Fiscal Year 2018: Support for Establishing Inter-University Exchange Programs with the United States and Other Countries Using the COIL Model]. https://www.jsps.go.jp/j-tenkairyoku/sentei_jigyo_h30.html (accessed May 1, 2025).

Japan Society for the Promotion of Science. (2018b). *Daigaku no Sekai Tenkairyoku kyouka jigyou (in Japanese)*" [INTER-UNIVERSITY EXCHANGE PROJECT] https://www.jsps.go.jp/english/e-tenkairyoku/index.html). Accessed November 25, 2025.

Japan Society for the Promotion of Science. (2026a). *Daigaku no Sekai Tenkairyoku Kyōka Jigyō(in Japanese).* https://www.jsps.go.jp/file/storage/j-tenkairyoku/data/sentei_jigyo/r5/r5tenkai_chousho_aa10.pdf

Japan Society for the Promotion of Science. (2026b). *Daigaku no Sekai Tenkairyoku Kyōka Jigyō(in Japanese).* https://www.jsps.go.jp/file/storage/j-tenkairyoku/data/sentei_jigyo/r5/r5tenkai_chousho_aa03.pdf

Japan Society for the Promotion of Science. (2026c). *Daigaku no Sekai Tenkairyoku Kyōka Jigyō (in Japanese).* https://www.jsps.go.jp/file/storage/j-tenkairyoku/data/sentei_jigyo/r5/r5tenkai_chousho_aa02.pdf

Ministry of Education, Culture, Sports, Science and Technology. (2018). *Sekai tenkairyoku jigyou (in Japanese)1-115.* [Inter- University Exchange Project (report)]. https://www.mext.go.jp/content/20220220-mxt_kotokoku01-000027716_1.pdf. *2.*

Ministry of Education, Culture, Sports, Science and Technology. (2025).*Daigaku no kokusaika no hitsuyosei (in Japanese*）[The Need for Internationalization in Universities] https://www.mext.go.jp/content/20241226-mxt_kotokoku02-000039459_7.pdf

Nagata, S. (2020). *Taiwa o unagasu kyoudo gakusyu: Eigo de Manabu akutibu laningu ni kansuru ichikosatsu (in Japanese*）[Active Learning Practices Using English as a Medium of Instruction in Japanese Higher Education]

O'Dowd, R. (2021). What do students learn in virtual exchange? A qualitative content analysis of learning outcomes across multiple exchanges. *International Journal of Educational Technology in Higher Education*, 18, 1–22. https://doi.org/10.1186/s41239-021-00250-1

Rubin, J., & Guth, S. (2016). Collaborative online international learning: An emerging format for internationalizing curricula. In R. O'Dowd & T. Lewis (Eds.), *Online intercultural exchange: Policy, pedagogy, practice* (pp. 15–27). Routledge.

The SUNYCOIL Center (2025). https://coil.suny.edu.

Blind Spots in “Language Support” for Multilingual Students in Japan

Kuznetsova Yulia

University of Tsukuba

s2530366@u.tsukuba.ac.jp

Abstract

Language support in Japan’s compulsory schools has expanded alongside growing linguistic diversity, yet in policy and research, the category is often treated as more stable and self-evident than it is. This article argues that language support is administratively useful but analytically unstable. Drawing on national guidance, local variation in student identification and support, and illustrative empirical studies, it develops a three-part typology of instability: epistemic, institutional, and subject-position. It shows that language support does not denote a single comparable category across contexts, because its meaning varies with local identification practices, institutional arrangements, and the positions assigned to learners. The article shows how these shifts affect what can, and cannot, be validly compared in research and policy.

Keywords

language support, language policy, educational categorization, multilingual students, Japan

1 Introduction

Across Japanese compulsory education, language support has grown in scope and prominence. Public schools enroll about 151,000 foreign-national students, and approximately 69,000 students are identified as needing Japanese-language instruction, including both non-Japanese students and Japanese students with international backgrounds (Ministry of Education, Culture, Sports, Science and Technology [MEXT], 2026a, pp. 1–2). Policy documents also highlight ongoing inequality in educational paths, especially at the high school level, where advancement and dropout rates remain notably uneven (MEXT, 2026a, p. 38). Together, these conditions have made language support a pressing issue in both policy and practice.

Yet despite its wide administrative use, language support has not been examined consistently as an analytic category in research on Japanese schooling. The term is used for pull-out Japanese-language instruction, support within or alongside mainstream classrooms, and, at transition points, special admissions or designated pathways for foreign-background students (MEXT, 2026a; Tokyo Gakugei University, 2023). Policy materials make clear that support does not follow a single national model: schools and municipalities use different procedures to identify students, and provision varies by local capacity, school type, and school

stage (MEXT, 2026a).

The problem, then, is not only uneven implementation but also interpretive comparability. When different identification practices, eligibility criteria, decision-making authorities, and forms of provision are collapsed into a single label, findings can appear more comparable than they really are. Differences in participation or outcomes may reflect not only learner characteristics or local provision, but also how support is defined and enacted in particular settings.

Recent policy discourse has increasingly acknowledged diversity through frameworks such as *tabunka kyōsei* ("multicultural coexistence") and revisions to the *gakushū shidō yōryō* (Course of Study) (MEXT, 2019, 2021). The policy materials reviewed here reflect this shift by differentiating foreign-background students in multiple ways, including Japanese proficiency, school stage, language background, region, migration history, and school-entry trajectory (MEXT, 2026a). Yet they continue to organize these learners within a broad support framework. Policy now recognizes greater internal diversity, but it does not settle what counts as support or how support should be compared across contexts.

This study, therefore, treats language support as administratively useful but analytically unstable. It develops a three-part typology of instability (epistemic, institutional, and subject-position) to show why support cannot be assumed to function uniformly across contexts. Its contribution is analytic rather than evaluative: it does not assess program effectiveness or propose an alternative model. Although the article centers on compulsory schooling, it also considers transition into upper secondary education where this reveals how earlier identification and support categories continue to shape access and recognition.

2 Analytic Stance and Materials

This article takes a category-analytic approach to language support. It examines how classificatory labels shape what becomes visible, measurable, and comparable in research and administration (e.g., Bowker & Star, 1999; Hacking, 2004). The study is a conceptual essay based on documentary and published sources rather than on new empirical fieldwork. It aims to clarify how the category of language support shapes who is counted as needing support, what kinds of provision are grouped together under the same label, and what can validly be compared across contexts.

This approach is important in Japan, where policy materials suggest that demographic change has not been matched by a fully shared identification framework. Some schools rely on classroom observation, while others use locally standardized assessments, ad hoc checklists, or consultation with external coordinators (MEXT, 2026a). The materials reviewed here do not indicate a fully standardized national protocol for identification across contexts. As a result, multilingual students with similar linguistic profiles may be categorized differently from one site to another, and eligibility may reflect local administrative practice as much as learner characteristics.

Policy materials further show that identification and provision do not follow a single standardized model. Students may be identified through teacher observation, check sheets, assessment informed by the Dialogic

Language Assessment (DLA), locally developed tools shared through platforms such as Casta-net, and the involvement of coordinators or other support personnel. The same materials also show that both the target population and the forms of support are broader than they may first appear. The population includes not only foreign-national students but also some Japanese-nationality students and, in some cases, children whose enrollment status remains uncertain. Support extends beyond Japanese-language instruction to include mother-tongue support, transition guidance, family liaison, and multiple instructional models, including pull-out, in-class, and classroom-integrated forms of learning (MEXT, 2026b, 2026c, 2026d).

Rather than treating language support as a transparent descriptor of learner populations, this analysis approaches it as an object of inquiry. It draws on three types of materials: (i) MEXT guidance documents and recent policy discussions concerning the education of foreign students; (ii) illustrative cases of local variation in student identification and support, including differences in eligibility criteria, assessment practices, and forms of provision; and (iii) selected empirical studies in which language support shapes participant inclusion and research interpretation.

These materials include broad policy discussions and more operational guidance on student acceptance, language assessment, and upper-secondary access. They are used here not as direct descriptions of classroom life, but as official statements that shape expectations and responsibilities in practice. The local cases serve as illustrative contrasts rather than nationally representative samples. The empirical studies show how language support functions analytically in research by shaping who is included, compared, and interpreted. Taken together, these materials make it possible to trace how the meaning of language support shifts across administrative and research contexts.

3 A Typology of Instability in Language Support

In Japanese schooling, language support commonly refers to Japanese as a Second Language (JSL) instruction, pull-out provision, and related eligibility categories for students judged to have limited Japanese proficiency. Administratively, the term may appear straightforward, but analytically it is unstable because it depends on multiple criteria that are only partly aligned across sites and institutions. These criteria shape how learners are identified, supported, and positioned in schools. This section develops the typology introduced earlier:

- epistemic instability: how “language need” is defined and measured;
- institutional instability: who has the authority to decide eligibility;
- subject-position instability: how eligibility constructs who the learner is taken to be.

3.1 Epistemic Instability

Epistemic instability refers to the absence of a shared basis for deciding which students need language support. In principle, eligibility is linked to Japanese-language proficiency. In practice, however, schools and

municipalities rely on different kinds of evidence, and no fully standardized national protocol governs identification across contexts (MEXT, 2026a).

Recent policy materials make this variation explicit. Schools may identify students through teacher observation, check sheets, DLA-informed assessment, locally developed tools, or judgments made by coordinators, rather than through a single nationally shared procedure (MEXT, 2026b, 2026c). The same materials also distinguish between surface-level Japanese and the deeper language demands of academic participation. In this sense, assessment is presented not only as a way to decide placement, but also as a basis for planning instruction (MEXT, 2026b, 2026c). Policy discussions further note that schools do not always have clear ways to distinguish Japanese-language acquisition needs from disability or special educational needs, which can complicate decisions about what kind of support a student requires (MEXT, 2026d).

Research on the development of the DLA likewise shows that assessment in this area is neither simple nor fully standardized. Butler and Sakurai (2020) show that the DLA was designed as a formative, dialogic tool that must take account of factors such as text difficulty, kanji density, teacher scaffolding, and the gap between oral proficiency and literacy development. At the same time, they note that assessment remains difficult to align fully with mainstream classroom demands and high-stakes entrance examinations. This suggests that judgments about language-related needs do not simply record an already obvious condition; they are shaped by the design of assessment tools, the interactional support provided during assessment, and the institutional context in which results are used. As a result, students with similar linguistic repertoires may be categorized differently across municipalities, complicating research comparison and administrative reporting.

3.2 Institutional Instability

Institutional instability means that schools and local authorities do not organize and provide language support in the same way. Without a single national framework, support is arranged differently across municipalities, boards of education, and schools, so the same category may function differently from one institution to another.

Substantial regional variation in access to Japanese-language instruction reflects not only learner need but also differences in local capacity (Burgess, 2011). Some areas have established JSL programs, while others depend on temporary funding, volunteers, or community partnerships. As a result, both eligibility thresholds and forms of provision vary widely. Policy discussions also distinguish between concentrated areas, where experienced personnel are more available, and dispersed areas, where assessment and support are often less stable because specialist staff are scarce (MEXT, 2026c, 2026e). Although MEXT now provides more guidance, reference frameworks, and advisor support, implementation still depends heavily on local boards of education, school leaders, coordinators, and teachers responding to regional conditions (MEXT, 2026c, 2026d). Local governments also use different delivery models: base-school systems, visiting instruction, and ICT-supported remote teaching (MEXT, 2026b, 2026c).

Locally specific arrangements make this visible. Prefectural arrangements such as Osaka's *waku-kō* (designated high schools with special admissions and support tracks for foreign-background students) and international-elective offerings in Kanagawa show how access and recognition are shaped through locally specific institutional responses (Tokunaga, 2017; Yamamoto & Enoi, 2023). Kobayashi and Tsuboya (2021) similarly argue that local access arrangements can create important pathways, but these remain dependent on local capacity rather than nationally shared guarantees. Policy discussions also identify transition to high school as a weak point in continuity of support, where provision often becomes thinner even though risks of dropout and exclusion increase (MEXT, 2026a). Access, therefore, depends not only on learner needs but also on how local systems are organized and resourced. The category does not function as a consistent institutional entity nationwide.

3.3 Subject-Position Instability

Subject-position instability means that language support does not only provide help; it also shapes how students are seen and what kinds of participation are expected from them, in line with sociocultural views of identity as positioning (Bucholtz & Hall, 2005). These positions are shaped in part by the difference between *kokugo* (Japanese as the national language) and *nihongo* (Japanese for non-native learners). This difference creates different expectations about who counts as a legitimate speaker and how students should participate in class (Tanaka, 2018). Students who receive language support are often seen as learners who need to move toward mainstream norms, rather than as multilingual students whose full linguistic resources belong in the classroom.

Recent policy discourse also frames these learners in more positive terms: as multilingual, capable, and rights-bearing students whose backgrounds should be treated as strengths rather than deficits (MEXT, 2026c, 2026e). Policy discussions emphasize diversity as an asset, encourage the use of mother tongue knowledge and prior learning, and suggest that the majority environment must also change in order to support coexistence (MEXT, 2026c). Yet this shift in framing does not fully resolve how students are positioned in practice.

Ethnographic research shows a recurring dilemma. Learners may remain in mainstream classrooms that they cannot fully access, or they may be pulled out to support classes that emphasize basic communication with limited academic content (Kanno, 2008). Research on JSL provision similarly shows that support spaces may function as places of safety and belonging while remaining weakly connected to age-appropriate academic content and mainstream curricular recognition (Moorehead, 2013; McGuire & Tokunaga, 2020). These arrangements, though intended to assist, can reinforce conditional membership. This contrasts with *kikokushijō* (returnee students), whose bilingualism is more often framed as an asset and supported through special admission pathways (Kanno, 2003). This positioning is reinforced when students' heritage languages and broader multilingual repertoires remain peripheral to ordinary curricula, signaling that they are not central to legitimate academic participation. The result is a tension between policy language that values diversity and

institutional arrangements that continue to organize students mainly through Japanese acquisition and adjustment.

4 Consequences of Instability for Research, Practice, and Experience

This section examines the consequences of category instability for how research is interpreted and how support is organized and experienced in schools. These should be read as category effects rather than as judgments about program quality or teacher intention. The discussion below focuses on the interpretive limits this creates and on how ambiguity is experienced in practice, rather than on the outcomes of any single program.

4.1 Consequences for Research

First, epistemic instability makes comparison across studies difficult. Because eligibility for language support is established through locally variable forms of evidence rather than a single nationally standardized protocol (MEXT, 2026a, 2026b, 2026c), students grouped together under the label “receiving language support” in one study may not be the same kind of population as those grouped under the same label in another. Apparent differences in outcomes may, therefore, reflect differences in identification practices as much as program effects or learner characteristics.

Second, institutional instability makes research samples harder to compare. Responsibility for identification and provision is distributed across municipalities, boards of education, and individual schools, so access depends on local governance capacity rather than shared entitlements (Burgess, 2011; MEXT, 2026a, 2026d). At the same time, policy materials show that language support covers a wider range of students and support arrangements than the label may suggest, including some Japanese-nationality students, children whose enrollment status remains uncertain, mother-tongue support, transition guidance, and multiple instructional models (MEXT, 2026a). Studies drawing on quota-based access schemes, special admission pathways, or locally organized support may, therefore, be examining institutionally specific rather than nationally comparable populations (Kobayashi & Tsuboya, 2021).

Third, subject-position instability complicates interpretation because eligibility does not signify the same social position in all contexts. In some settings, support may function as temporary linguistic assistance; in others, it may signal more durable academic marginality, affective compensation without curricular inclusion, or a conditional pathway toward mainstream participation (Kanno, 2003, 2008; Moorehead, 2013; McGuire & Tokunaga, 2020). Treating language support as a neutral descriptor risks collapsing these different positions into a single category. Therefore, interpretation depends not only on who receives support, but also on what support means in a given institutional setting.

Taken together, these patterns suggest that findings in language-support research should often be read as category effects as much as program effects. Instability limits what can be validly compared across studies and narrows the scope of the claims such research can make.

4.2 Consequences for Practice and Experience

4.2.1 Uneven Access to Support

Where identification relies on different kinds of evidence and responsibility is spread across local institutions, access to support becomes uneven. In practice, support depends not only on student need but also on local expertise, staffing, assessment capacity, and the way support is organized. Policy discussions explicitly identify disparities between concentrated and dispersed areas, and they also note that support often weakens at transition points, especially in the move to high school (MEXT, 2026c, 2026d). This means that access is shaped not only by student characteristics but also by the local conditions under which support is organized.

Longitudinal work on Brazilian Nikkei students in Japan similarly shows that marginalization can persist despite changing policy discourse and expanding support initiatives (Hachiman & De Mello, 2025). Quantitative research on Brazilian migrant children also shows that local context matters for high school enrollment: residence in Hamamatsu, where integration programs and NPO support were more developed, was associated with higher odds of enrollment than in less supportive settings (Takenoshita et al., 2014). Large-scale quantitative research further shows persistent disparities in high school enrollment between native and immigrant-background children, even when some socioeconomic differences are considered (Hagiwara & Liu, 2023). Together, these studies suggest that expanding support and provision does not by itself produce stable or equitable access.

4.2.2 Peripheralization of Multilingual Repertoires

A second consequence is that students' multilingual repertoires may remain peripheral to formal schooling. Even where policy discourse increasingly values diversity, support often continues to treat students' non-dominant languages as secondary to the main task of acquiring Japanese. These languages may be welcomed socially or used instrumentally, but they are often weakly connected to age-appropriate curriculum, assessment, and formal recognition.

Research on JSL provision illustrates this pattern clearly. Moorehead (2013) shows that support spaces may function as sites of curricular suspension, weakly connected to mainstream academic content and marginal to core school routines. McGuire and Tokunaga (2020) similarly show that support classrooms can function as ibasho, spaces of safety and belonging, while mainstream classrooms remain governed by fixed curricular expectations and assimilative norms. In such arrangements, students' broader linguistic repertoires may be socially affirmed without becoming academically central. The result is a form of inclusion in which multilingualism is acknowledged, but not fully recognized as part of legitimate classroom participation.

4.2.3 Support Without Institutional Legitimacy

A third consequence is that support may exist without being fully institutionalized within ordinary schooling. When eligibility is interpreted differently across settings and provision depends heavily on local organization, support may be available, but its status and purpose remain unstable (Burgess, 2011; MEXT, 2026a).

Kobayashi and Tsuboya (2021) show that local access arrangements can create important pathways into high school. However, these pathways often operate within unchanged school expectations and standards of evaluation. In other words, support may help students to enter or to stay in school without changing how they are judged once they are there. Under these conditions, support can function as help or accommodation without giving students stable academic legitimacy. As a result, uncertainty about what support is meant to do becomes part of everyday schooling.

5 Why Instability Persists: Ideological and Policy Conditions

The instabilities identified above do not persist simply because schools lack better tests, clearer procedures, or stronger coordination. They persist because current arrangements allow schools to respond to linguistic diversity without substantially changing the main norms of participation, evaluation, and curriculum. This does not mean that instability is produced intentionally. Rather, the system can continue to function without fully resolving it. This section examines the ideological and policy conditions that help to sustain instability in the category of language support.

5.1 Monolingual Norms in Mainstream Schooling

A key reason why instability persists is that compulsory schooling remains organized around monolingual norms. Instruction and assessment remain organized around *kokugo* as the unmarked medium of legitimate participation, while *nihongo* functions as a supplementary domain for students classified as non-native or non-standard users (Heinrich, 2012; Sato, 2019). This distinction may expand access to support, but it also preserves a hierarchy in which mainstream participation continues to be judged through Japanese-centered expectations.

Within this framework, policy may acknowledge diversity, but everyday classroom participation is still organized around Japanese as the main legitimate medium of learning. Recent policy materials describe diversity in more positive terms and encourage schools to draw on students' prior knowledge and mother-tongue resources (MEXT, 2026b, 2026c, 2026e). Even so, these shifts do not by themselves change how participation is recognized in mainstream instruction and assessment. Language-support categories can, therefore, expand without fundamentally changing *kokugo*-centered norms. Thus, instability reflects not only administrative variation but also a continuing tension between multilingual realities and monolingual institutional expectations (Kubota & McKay, 2009).

5.2 Hierarchies of Valued Multilingualism

Instability also persists because not all forms of multilingualism are equally recognized in school. In policy and curriculum discourse, "foreign language education" remains closely associated with English, while other linguistic repertoires occupy a much weaker position in curriculum design, teacher preparation, and assessment

(Honna & Saruhashi, 2019). Languages other than English often lack standardized pathways, stable funding, and widely available teaching materials, leaving support for broader multilingual development uneven and locally dependent (Mizuguchi & Hasegawa, 2016).

This hierarchy affects not only which languages are taught, but also which languages become visible as legitimate school knowledge. Heritage and other non-dominant languages may be valued symbolically or privately, yet they are seldom treated as central to academic participation. As a result, many students rarely encounter these languages as recognized parts of the formal curriculum. Policy materials increasingly frame students' diversity as a strength and refer positively to mother-tongue knowledge and multilingual resources (MEXT, 2026b, 2026c, 2026e). At the same time, these resources remain weakly institutionalized in mainstream schooling. Support, therefore, tends to focus on helping students participate through Japanese, while other parts of their linguistic repertoires remain supplementary or symbolic. Under these conditions, ambiguity surrounding language support comes to seem normal rather than unusual, because the wider hierarchy of valued languages remains largely unchanged.

5.3 Institutional Reproduction of Ambiguity

These ideological and policy hierarchies are reproduced through the way support is organized in practice. Even as national guidance has expanded, implementation still depends heavily on local expertise, staffing, coordination, and continuity (MEXT, 2026a). As a result, support develops unevenly across settings, not only because needs differ, but because the capacity to respond to them is unevenly distributed.

For this reason, ambiguity is reproduced not only through policy gaps, but through ordinary institutional arrangements that remain uneven from place to place. Support may exist, but it is often only loosely connected to mainstream instruction, assessment, and long-term progression. Studies report social isolation, bullying, emotional distress, and language-related difficulties, particularly among later-arriving students (Inoue & Banstola, 2025). Research also shows that support spaces may provide safety or belonging while remaining weakly connected to mainstream academic recognition (Moorehead, 2013; McGuire & Tokunaga, 2020). These studies do not explain instability on their own, but they show how it is sustained in practice through arrangements that provide assistance without fully reorganizing participation, assessment, or curricular legitimacy.

Taken together, these patterns suggest that instability in language-support categories is not simply a technical problem. Schools and systems may expand support while leaving broader norms of recognition largely unchanged. This helps to explain why ambiguity persists even as policy discourse becomes more inclusive.

6 Conclusion

This study has argued that *language support* in Japanese schooling is useful as an administrative label, but unstable as an analytic one. The problem is not only uneven implementation. Schools identify students in different ways, different institutions decide eligibility, and support positions learners differently across settings. As a result, the term does not always refer to the same kind of student group, provision, or institutional role.

The typology developed in this study, epistemic, institutional, and subject-position instability, shows how this variation is produced through ordinary school and research practices. For research, it makes studies harder to compare because students grouped under this label in one setting may not be equivalent to those grouped under it elsewhere. In practice, it is associated with uneven access, the marginal place of students' broader multilingual repertoires, and forms of support whose place within ordinary schooling remains unclear. These are best understood as effects of how the category is constructed and applied, rather than simply as evidence that a particular program or teacher is effective or ineffective.

One implication is that differences in achievement, participation, or support outcomes may reflect differences in how the category is defined and organized, not only differences in students or teaching. The point is not to argue against support or to suggest that instability can be removed completely. Instead, it is to show that support can expand while the main norms of recognition in schooling remain largely unchanged. As long as schools and local systems define language support differently, conclusions drawn from that category need to specify who counted as receiving support, how that decision was made, and what the support actually involved.

References

Bowker, G. C., & Star, S. L. (1999). *Sorting things out: Classification and its consequences.* MIT Press.

Bucholtz, M., & Hall, K. (2005). Identity and interaction: A sociocultural linguistic approach. *Discourse Studies, 7*(4–5), 585–614. https://doi.org/10.1177/1461445605054407

Burgess, C. (2011). (Mis) Managing diversity in non-metropolitan public schools: The lack of state-sponsored support for "newcomer" children. In R. Tsuneyoshi, K. H. Okano, & S. Boocock (Eds.), *Minorities and education in multicultural Japan: An interactive perspective* (pp. 191–212). Routledge.

Butler, Y. G., & Sakurai, C. (2020). Developing a classroom-based language assessment of Japanese for children who speak minority languages in Japan: The dialogic language assessment. *Language Assessment Quarterly, 17*(5), 467–490. https://doi.org/10.1080/15434303.2020.1826487

Hachiman, E., & De Mello, R. (2025). Escolaridade de crianças e adolescentes brasileiros no Japão: Desafios migrantes na atualidade. *Educação & Sociedade, 46*, e289099. https://doi.org/10.1590/es.289099

Hacking, I. (2004). Between Michel Foucault and Erving Goffman: Between discourse in the abstract and face-to-face interaction. *Economy and Society, 33*(3), 277–302. https://doi.org/10.1080/0308514042000225671

Heinrich, P. (2012). *The making of monolingual Japan: Language ideology and Japanese modernity*. Multilingual Matters.

Honna, N., & Saruhashi, J. (2019). Language education policy in Japan. In A. Kirkpatrick & A. J. Liddicoat (Eds.), *The Routledge international handbook of language education policy in Asia* (pp. 97–110). Routledge. https://doi.org/10.4324/9781315666235

Inoue, S., & Banstola, R. S. (2025). Adjustment and adaptation issues affecting the social, emotional and psychological well-being of immigrant children in Japanese schools. *Sage Open, 15*(2). https://doi.org/10.1177/21582440251340627

Kanno, Y. (2003). *Negotiating bilingual and bicultural identities: Japanese returnees betwixt two worlds*. Lawrence Erlbaum Associates.

Kanno, Y. (2008). *Language and education in Japan: Unequal access to bilingualism*. Palgrave Macmillan.

Kobayashi, H., & Tsuboya, M. (2021). Social resources and challenges related to the schooling and education of immigrant children at high schools in Japan. *International Migration & Integration, 22*, 369–384. https://doi.org/10.1007/s12134-019-00752-2

Kubota, R., & McKay, S. (2009). Globalization and language learning in rural Japan: The role of English in the local linguistic ecology. *TESOL Quarterly, 43*(4), 593–619. https://doi.org/10.1002/j.1545-7249.2009.tb00188.x

McGuire, J. M., & Tokunaga, T. (2020). Co-constructing belonging: "Voluntary separation" in deaf and immigrant education in Japan. *Japanese Studies, 40*(3), 291–311. https://doi.org/10.1080/10371397.2020.1851177

Ministry of Education, Culture, Sports, Science and Technology (MEXT). (2019). *Gaikokujin no ukeire · kyōsei ni muketa gaiyō* [Outline of initiatives for acceptance of foreigners and coexistence]. https://www.moj.go.jp/isa/content/930004524.pdf

Ministry of Education, Culture, Sports, Science and Technology (MEXT). (2021). *Reiwa no Nihon-gata gakkō kyōiku no kōchiku o mezashite: Subete no kodomo-tachi no kanōsei o hikidashisu, kobetsu saiteki na manabi to, kyōdōteki na manabi no jitsugen (tōshin)* [Toward building the Reiwa-era model of Japanese-style school education: Realizing individually optimized learning that draws out every child's potential and collaborative learning (Advisory report)]. https://www.mext.go.jp/content/20210524-mxt_kyokoku-000015284_12.pdf

Ministry of Education, Culture, Sports, Science and Technology (MEXT). (2026a). *Gaikokujin jidō seito tō no kyōiku ni kansuru genjō to kadai ni tsuite* [Current status and issues regarding the education of foreign children and students] [Report]. International Education Division, Comprehensive Education Policy Bureau. https://www.mext.go.jp/content/20260220-mxt_kyokoku-000047375_006.pdf (in Japanese)

Ministry of Education, Culture, Sports, Science and Technology (MEXT). (2026b). *Shidō naiyō no shinka・jūjitsu ni kansuru kore made no giron to taiō no hōkōsei* [Summary of discussions to date and direction for deepening/enhancing instructional content] [Meeting handout]. https://www.mext.go.jp/content/20260220-mxt_kyokoku-000047375_01.pdf

Ministry of Education, Culture, Sports, Science and Technology (MEXT). (2026c). *Gaikokujin jidō seito tō no kyōiku no jūjitsu ni kansuru yūshikisha kaigi hōkokusho kosshi (an)* [Draft outline of the Expert Meeting report on enhancing education for foreign children and students] [Meeting handout]. https://www.mext.go.jp/content/20260220-mxt_kyokoku-000047375_02.pdf

Ministry of Education, Culture, Sports, Science and Technology (MEXT). (2026d). *Gaikokujin jidō seito tō no kyōiku no jūjitsu ni kansuru yūshikisha kaigi ni okeru kore made no giron no seiri* [Organizing the discussions to date in the Expert Meeting on enhancing education for foreign children and students] [Meeting handout]. https://www.mext.go.jp/content/20260220-mxt_kyokoku-000047375_03.pdf

Ministry of Education, Culture, Sports, Science and Technology (MEXT). (2026e). *Gaikokujin jidō seito tō no kyōiku no jūjitsu ni kansuru yūshikisha kaigi ni okeru omona kentō jikō* [Main items considered in the Expert Meeting on enhancing education for foreign children and students] [Meeting handout]. https://www.mext.go.jp/content/20260220-mxt_kyokoku-000047375_003.pdf

Mizuguchi, K., & Hasegawa, Y. (2016). Kōtō gakko no tagengo kyōiku no genjō – Seisaku no hinkon to genba no doryoku – [The current status of multilingual education in high schools – Poor policies and efforts made from educational side]. In M. Morizumi, A. Furuishi, M. Sugitani, & Y. Hasegawa (Eds.), *Gaikokugo kyōiku wa eigo dake de iinoka: Global shaki wa tagengoda!* [Is it good to have only English in foreign language education: Global society is multilingual!] (pp. 172–189). Kuroshio Shuppan.

Moorehead, R. (2013). Separate and unequal: The remedial Japanese language classroom as an ethnic project. *Asia-Pacific Journal, 11*(32), e2. https://doi.org/10.1017/S1557466013034499

Sato, G. (2019). *Tabunka shakai ni ikiru kodomo no kyōiku: Gaikokujin no kodomo, kaigai de manabu kodomo no genjō to kadai* [Education of children living in multicultural society: The current situation and tasks for foreign children and children studying overseas]. Akashi Shoten.

Takenoshita, H., Chitose, Y., Ikegami, S., & Ishikawa, E. A. (2014). Segmented assimilation, transnationalism, and educational attainment of Brazilian migrant children in Japan. *International Migration, 52*(2), 84–99. https://doi.org/10.1111/imig.12057

Tanaka, L. (2018). Japanese language research: Kokugo as an ideology, nihongo as an autonomous and global scholarship? In K. Okano & Y. Sugimoto (Eds.), *Rethinking Japanese studies: Eurocentrism and the Asia-Pacific region* (pp. 32–52). Routledge.

Tokunaga, T. (2017). Multicultural education in Japan. In *Oxford Research Encyclopedia of Education*. https://doi.org/10.1093/acrefore/9780190264093.013.65

Tokyo Gakugei University. (2023). *Kōtō gakkō ni okeru gaikokujin seito-tō no ukeire no tebiki* [Guidelines for the acceptance of students with foreign roots in upper secondary education]. https://www2.u-gakugei.ac.jp/~knihongo/feature/upload/koko_nihongo_tebiki.pdf

Yamamoto, K., & Enoi, Y. (Eds.). (2023). *Gaikokujin seito to tomo ni ayumu Ōsaka no kōkō — Gakkō bunka no hen'yō to sotsugyōsei no raifu kōsu* [High schools in Osaka working with foreign-background students: Transforming school culture and the life courses of graduates]. Akashi Shoten.

English Education and Part-Time Lecturers in Japanese Universities: A Case-Based Analysis

Yukiko Ideno[1] , Sumie Akutsu[2] , Shin Sakuragi[3], and Sachiko Takagi[4]

[1]Tokiwa University, [2]Toyo University, [3]Shibaura Institute of Technology, [4]Tokiwa University

mlg19817@nifty.com, akutsu@toyo.jp, sakuragi@shibaura-it.ac.jp, takagi@tokiwa.ac.jp

Abstract

This study investigates the challenges faced by part-time English lecturers at universities in Japan and suggests ways to improve English courses as a part of university-level language education. Japanese higher education, especially private universities, heavily relies on adjunct lecturers to deliver curricula for its English education. However, since the late 1990s, social changes have made it increasingly difficult for Japanese universities to hire qualified part-time English instructors, threatening the quality and sustainability of English language education. This study focuses on two faculties at two private universities in Tokyo, where over 80% of English classes are taught by part-time lecturers. By comparing faculties with different majors, it aims to identify both shared and unique issues among part-time instructors. A survey with open-ended questions was conducted among current and former adjunct English lecturers, covering the following themes: teaching-related concerns, coping strategies, maintaining sustainable teaching, supporting students before COVID-19, supporting students during and after COVID-19, and applying online teaching experience in current instruction. The data were analyzed using KH Coder, a text-mining software, to extract both positive and negative patterns. Results indicate that respondents often experience issues such as unclear job responsibilities, insufficient institutional support, and job insecurity. Nevertheless, they make strong efforts to facilitate student learning through quizzes, group work, and online tools. The findings suggest that how instructors cope with challenges depends on their individual backgrounds and the level of support provided by administrators of the courses. Based on these results, we argue that it is crucial for course administrators to offer clearer job descriptions, access to more teaching materials, and stronger institutional support to part-time English instructors. The study concludes that collaboration among universities through the sharing of challenges and solutions can provide valuable insights into improving English education in Japan.

Keywords

English Education, Part-time Lecturers, Sustainable Education

1 Introduction

Japanese universities, especially large private ones, have been heavily dependent on part-time instructors, and English education is no exception. The number of part-time English instructors has been steadily increasing as the number of universities grows. This reliance has become a defining feature of higher education nationwide. This situation may be justified from a management perspective. The COVID-19 pandemic further exposed these issues, particularly for part-time instructors who often have limited access to digital tools and training opportunities.

Previous research has highlighted the significant role of part-time instructors in English education at Japanese universities (Kato, 2023). Despite unstable employment conditions and limited access to institutional resources (Kanbayashi, 2021), many of these instructors continue to deliver effective instruction. According to Japan's Ministry of Education, Culture, Sports, Science and Technology (MEXT), their adaptability and creativity during and after the COVID-19 pandemic have been particularly notable (MEXT, 2022). They learned to integrate online tools and hybrid teaching styles, enabling them to sustain classes even under difficult circumstances. This experience also transformed perceptions of communication and feedback among both teachers and students (Yamaguchi, 2021). According to Taguchi (2022), shifts in student satisfaction data suggest that education now places greater emphasis on clarity, flexibility, and emotional understanding in teacher–student relationships. In other words, students increasingly value not only knowledge, but also how instructors listen, respond, and support them in diverse learning environments.

However, several critical issues remain unresolved. Part-time instructors frequently report heavy teaching loads and insufficient institutional support (Kimura, 2016; Kanbayashi, 2021). They often prepare materials, adjust lesson levels, and manage classes independently, highlighting the need for improved communication between universities and part-time faculty (Kato, 2023). As emphasized by MEXT (2020; 2022), sustainable education must consider both teaching quality and teacher well-being. Therefore, universities should try to support part-time instructors by improving their working environments, which in the long run would benefit both educators and students.

This study tries to gain insights into what kind of support universities should provide to part-time instructors to improve and sustain English education by examining their working environments and teaching experiences of part-time English instructors. It focuses on their teaching challenges, adaptations before and after COVID-19, and university efforts to promote sustainability and student satisfaction. Data were collected through a questionnaire survey of current and former part-time instructors, and both quantitative and qualitative responses were analyzed using text-mining and correspondence analysis. By identifying key concerns and innovative teaching approaches, this study aims to offer data-driven suggestions for developing a more sustainable English education system in Japan, particularly regarding the role and engagement of part-time teachers.

2 Method

2.1 Participants and Basic Information

A total of 23 valid responses were collected from part-time university English instructors. The participants included both current and former part-time English instructors who were recruited through professional networks and academic contacts. Among them, 14 were female, 7 were male, and 2 did not indicate their gender. Participants began their teaching careers between 1992 and 2022. The average number of courses taught per semester was 8.65 (*SD* = 5.31), and the average number of English courses taught was 5.22 (*SD* = 2.62).

2.2 Questionnaire

The questionnaire consisted of six open-ended items focusing on the teaching experience and challenges of part-time English instructors. This questionnaire was administered online using Google Forms and was written in Japanese. The items were as follows:

1) Concerns and Difficulties: Do you have any anxieties, concerns, or difficulties related to teaching English as a part-time instructor?

2) Coping Strategies: How do you deal with these concerns, or how would you like to address them in the future?

3) Sustainability and SDGs (Quality Education): In relation to the SDG goal of Quality Education, what efforts do you make to ensure sustainable teaching as a part-time instructor?

4) Improving Student Satisfaction (Pre-COVID): Before the COVID-19 pandemic, what strategies did you use to improve student satisfaction in your classes?

5) Improving Student Satisfaction (Post-COVID): Since the COVID-19 pandemic, what strategies have you used to improve student satisfaction?

6) Application of Online Teaching Experience: How do you think your online teaching experience during the pandemic will be applied to future classes?

2.3 Data Analysis

In this study, each individual response was treated as a unit of analysis in KH Coder. The text data were segmented based on each questionnaire item, and all responses were analyzed collectively within each category. The selection of forced extraction words was based on their relevance to current digital learning environments and frequently used educational tools in university English education.

For analytical purposes, the responses were labeled as follows:

- Concerns and difficulties
- Coping strategies

- Efforts to ensure sustainability
- Pre-COVID: Efforts to improve student satisfaction
- Post-COVID: Efforts to improve student satisfaction
- Application of experience to future teaching

To analyze the free-text responses objectively, text mining was conducted using KH Coder, a software tool for quantitative content analysis. This method was selected to reduce researcher bias and to ensure objectivity in interpreting qualitative data. During preprocessing, grammatical particles, auxiliary verbs, and other function words that do not carry independent meaning were excluded from the analysis. Only content words, including nouns, verbal nouns, adjectives, adverbs, and verbs, were retained. Responses that contained only "nothing in particular" were excluded from the dataset. Additionally, the following expressions were manually included as forced extraction words to capture digital and pedagogical features relevant to current university English teaching: *ACE, MyELT, online learning, online assignments, Google Classroom, Google Form, Microsoft Teams, face-to-face, PowerPoint slides, group work, pair work, generative AI, on-demand, cloud, and PowerPoint.*

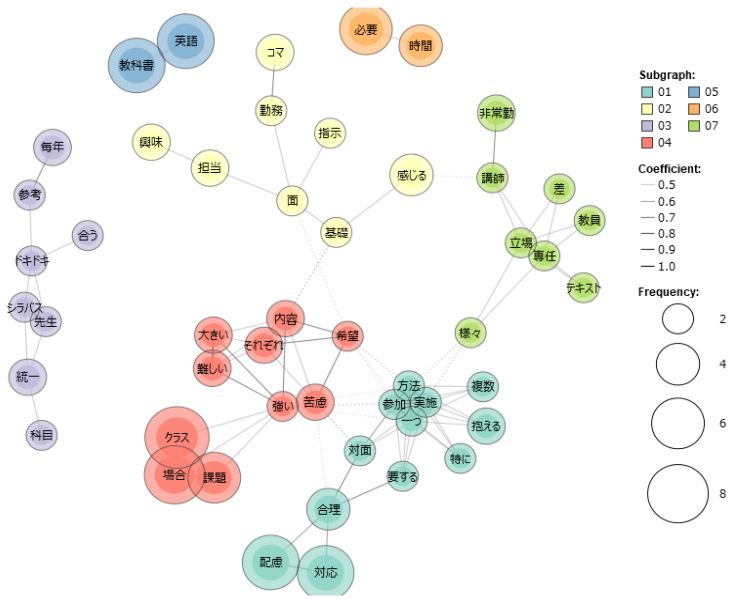
Through the integration of open-ended responses and text-mining methods, the study aimed to identify key terms, co-occurrence patterns, and conceptual clusters related to part-time instructors' concerns, strategies, and innovations. This mixed approach enabled an integrated understanding of how individual instructors' voices reflect broader issues of sustainability and quality in university-level English education.

3 Results and Analysis: Co-occurrence Network Analysis and Correspondence Analysis

3.1 Concerns and Difficulties in Language Teaching

Figure 1

Concerns and Difficulties in Language Teaching



In Figure 1, the cluster primarily seems to reflect emotional stress and pedagogical complexity. Teachers expressed that class content and student responses vary on a case-by-case basis, leading to difficulties in task-setting and emotional strain (Figure 1).

In this figure, the terms 苦慮 (struggle) and 難しい (difficult) are central, suggesting mental and emotional burdens related to classroom management. A small but

strong cluster emphasizes the foundational role of textbooks in English teaching. The nodes 英語 (English) and 教科書 (textbook) are isolated yet closely connected, indicating that issues surrounding textbook selection and use are consistently mentioned together, possibly reflecting a shared concern among respondents.

Another cluster reveals the precarious status of part-time language instructors. Words such as 立場 (position), 責任 (responsibility), 差 (gap) point to perceived inequalities and tensions between professional responsibility and institutional authority, indicating structural challenges related to employment conditions. The cluster containing 感じる (feel) and 指示 (order) seems to reflect the cognitive and emotional load experienced by instructors as they navigate both external expectations and internal workload, especially when managing multiple classes as indicated by コマ (unit). Although small, another cluster highlights a fundamental constraint suggested by 時間 (time) and 必要 (necessary), implying a lack of sufficient time to prepare adequately or respond to institutional demands, underscoring the stress caused by persistent time pressure.

One cluster centers on adaptive strategies, illustrating how teachers are experimenting with various teaching methods, balancing face-to-face and online teaching, and seeking practical solutions. Behind this is the revision of the Act for Eliminating Discrimination against Persons with Disabilities in 2021, which requires "reasonable accommodation" for disabled people. A smaller and peripheral cluster captures repetitive stressors and structural elements including シラバス (syllabus), ドキドキ (nervous), and 統一 (unification), which indicate ongoing anxieties about syllabus creation and calls for greater standardization.

Overall, the core tension appears to lie between emotional and pedagogical pressures and the adaptive strategies teachers employ. There is also a clear divide between institutional and employment concerns and classroom practices. Textbook-centered instruction and structural challenges emerge as peripheral yet persistent background issues.

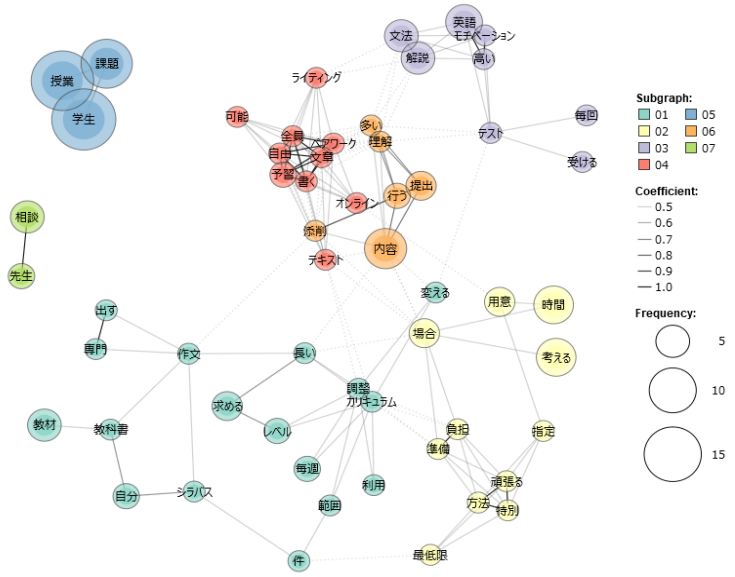
3.2 Concerns and Difficulties in Managing Teaching Tasks

Figure 2 appears to show a dense cluster centered on how instructors handle assignment design, student motivation, and classroom management challenges. This is particularly relevant to supporting and accommodating diverse learners as well as addressing their part-time status and irregular schedules. Terms such as 締切 (deadline) and 目標 (goal) highlight concerns with time-bound tasks, while モチベーション (motivation) may suggest reflection on student readiness and engagement (Figure 2).

Another cluster focuses on lesson content and delivery, indicating that lesson design is closely tied to assignment strategy and emphasizes flexibility in adapting content and teaching methods. Instructors also express concerns about balancing workload and time. The terms 適切な (appropriate) and 最低限 (minimum) imply efforts to maintain instructional quality while avoiding overburden. This cluster represents

Figure 2

Concerns and Difficulties in Managing Teaching Tasks



themes of time, workload, and flexibility, illustrating pragmatic adjustments in instructional expectations.

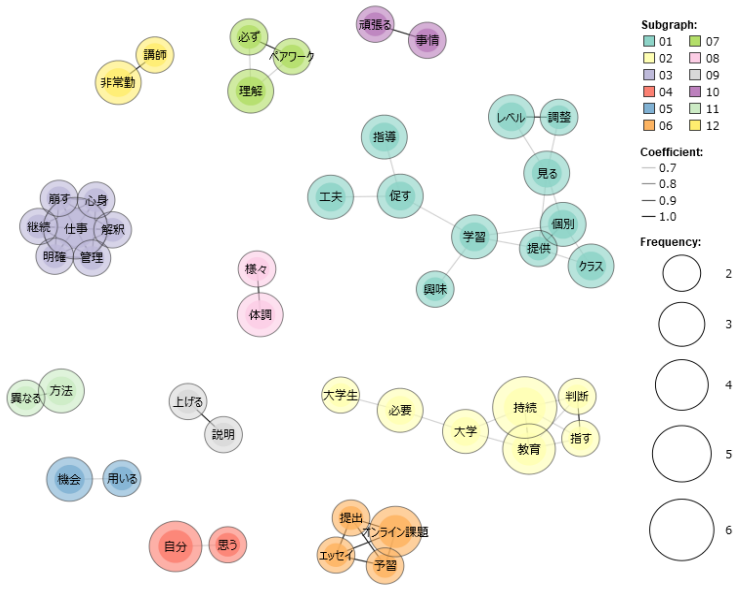
Although small, the isolated cluster containing 相談 (consultation) and 教員 (teacher) underscores the importance of mentorship or peer support, which may be underutilized or insufficient in some teaching contexts. Another cluster highlights practical aspects related to curriculum, materials, and weekly planning. Regarding course design, progression, and assessment, レベル (level) and アウトプット (output) indicate attempts to align assignments with students' abilities and learning goals. In connection with evaluation and assessment, the association among 感情 (emotion), 文法 (grammar), and テスト (test) appears to suggest the stress students experience around testing and the need for more thoughtful and supportive feedback. One particularly notable feature is the central triangle consisting of 授業 (class), 課題 (task) and 学生 (student), which represents the core teaching dynamic. These terms frequently appear together but remain somewhat isolated from other specific clusters. This may suggest that they are foundational but not deeply embedded in any single issue category.

These clusters reveal that instructors are actively managing multiple and competing priorities such as student diversity, task design, time constraints emotional load, and curricular coherence. There is a consistent emphasis on adaptation and balancing workload, reflecting shared concern about the sustainability of teaching practices. Moreover, course design and weekly structure appear central to implementation strategies, while assignment management and student engagement remain constant challenges for language instructors.

3.3 Ensuring Sustainability in Language Education

Figure 3 illustrates that sustainability and institutional commitment appear to be the central themes, revealing a growing awareness that institutional support is essential for maintaining the level of language education.

The close relationship between 持続 (continuation) and 判断 (Judgment) may suggest an ongoing need for clear direction and accountability at the university level. Another cluster demonstrates active efforts to personalize learning support strategies, as indicated by 個別 (individual), 学習 (learning), 指導



(instruction). There is a strong emphasis on individualized instruction, assessment-based adjustments, and student engagement, which may suggest that sustainability also depends on adaptability in teaching methods.

A simple but important cluster includes 理解 (comprehension) and チームワーク (teamwork), highlighting shared understanding and collaboration as key to long-term viability in teaching contexts, particularly for part-time instructors. Another group of clusters focuses on the use of online assignments and essay tasks to sustain engagement. These assignments serve as practical tools to maintain continuity, especially in post-COVID or hybrid learning environments. One unique and crucial cluster conveys a warning through the terms 心身 (mind-body), 壊れる (collapse), and 継続 (renewal). This underscores that inadequate support systems can lead to burnout, thereby undermining sustainability. There appears to be an urgent need for better workload management and emotional support. While practical concerns regarding tool access and opportunities for improvement are likely linked to previous online learning initiatives, one of the smallest clusters may suggest that Part-time instructors may be underrepresented within the educational system and may lack adequate institutional support. These isolated clusters reflect fragmented perspectives or underserved issues that require stronger integration into institutional discussions.

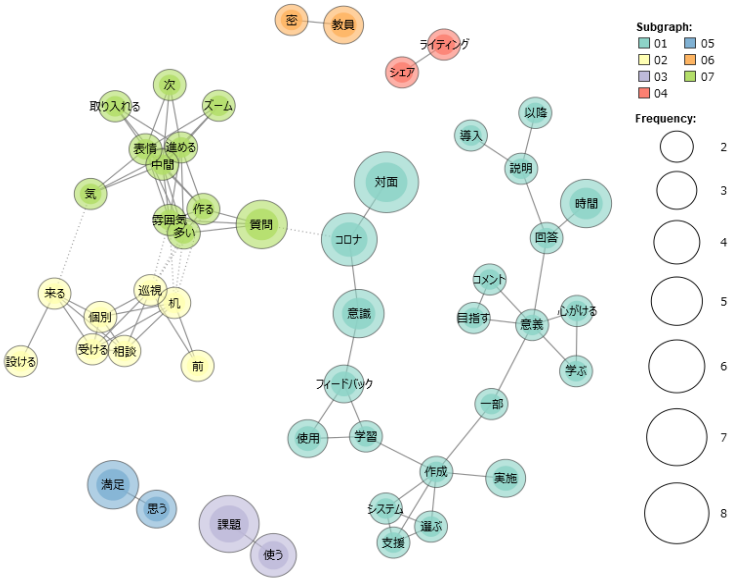
3.4 Enhancing Student Satisfaction (Pre-COVID)

With regard to the enhancement of student satisfaction before the COVID-19 pandemic, the cluster focuses on the institutional efforts to measure and implement various teaching styles (Figure 4).

The terms 活用 (utilization) and 得る (gain) emphasize feedback and reflective practices, particularly in comparing face-to-face and online learning experiences. Another cluster demonstrates student-oriented reflection. The terms such as 興味 (interest) and 意見 (opinion) suggest that learner-centered approaches were associated with higher levels of satisfaction prior to COVID-19. Many of the smaller clusters, including 学生 (student) and 授業 (class), seem to represent the primary beneficiary group.

Another group of clusters containing 指導 (instruction), 方法 (method), 英語 (English), 個別 (individual), 教育 (education), ニュース（news）, 映画（movies） seems to reflect concern about how

Figure 4



English language education was delivered, particularly the role of personalized instruction. This implies a demand for diversified pedagogical approach tailored to student needs. Emphasizing the role of instructors, one cluster including 発言 (remark) suggest that active involvement and verbal contributions by instructors were factors related to perceived satisfaction. Some other isolated terms, such as オンライン(online), 合わせる(combine), 聞く(listen), and 答える(answer), appear disconnected from the central themes, implying that technological components and assessment mechanics were not yet fully integrated into the discussion of student satisfaction prior to the pandemic.

Pre-COVID efforts to enhance student satisfaction appear to have emphasized learner-centered activities, reflective practices, hybrid instructional methods, and personalized teaching in language education. However, the limited integration of students' perspectives and online elements highlights areas for future development and stronger institutional alignment.

3.5 Enhancing Student Satisfaction (Post-COVID)

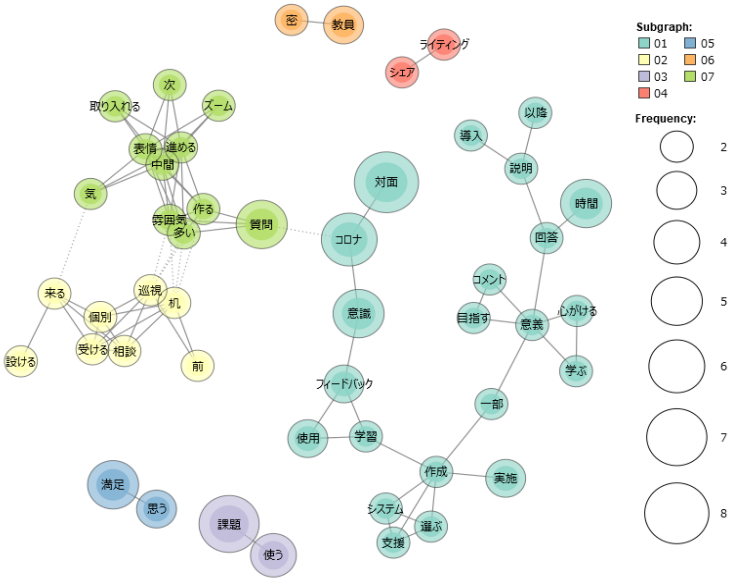
Compared to the pre-COVID19 results, the cluster reflecting the post-COVID educational landscape emphasizes blended learning, feedback mechanisms, and systematic support (Figure 5).

There is a strong pedagogical focus on interactivity, as shown by the terms 意識 (consciousness), フィードバック (feedback), and 作成 (creation), which reflect a learner-centered approach. The introduction of システム (system) and 支援 (support) further may suggest institutional adaptation aimed at sustaining student satisfaction.

Another cluster centers on the importance of personalized interaction, represented by 個別 (individual), 相談 (support), and オフィスアワー (office hour), which appear to have been encouraged or strengthened by remote and hybrid teaching contexts. Faculty responsiveness and individualized guidance play a key role in shaping the post-pandemic student experience. One cluster highlights the emotional climate of post-COVID classrooms regarding communication. The frequent co-occurrence of terms related to Zoom and emotional cues such as 表情 (facial expression) and 雰囲気 (atmosphere) seems to reflect

Figure 5

Enhancing Student Satisfaction (Post-COVID)



the hybrid learning environment and its demand for non-verbal affective engagement. Another area featuring 満足 (satisfaction) may suggest student self-reflection on the effectiveness of digital tools and instructional delivery, pointing to evolving perceptions of learning quality. A smaller and technology-oriented cluster points to collaborative or asynchronous activities such as ライティング (writing) and 共有する (share). It is also noteworthy that one small cluster implies lingering concerns about COVID-related safety protocols and teacher roles in managing them, as suggested by the term 密 (crowded space).

Compared with the pre-COVID environment, there is a stronger emphasis on feedback systems, digital tool utilization, and hybrid instruction methods, all of which have become more prevalent in the post-COVID teaching context. Moreover, emotional awareness, and sustained engagement mechanisms have emerged as key drivers of student satisfaction. Finally, the continued appearance of terms such as closed spaces and Zoom indicate a lasting sensitivity to pandemic-related dynamics, even as face-to-face instruction resumes.

3.6 Co-occurrence Network Analysis: Application of Experience

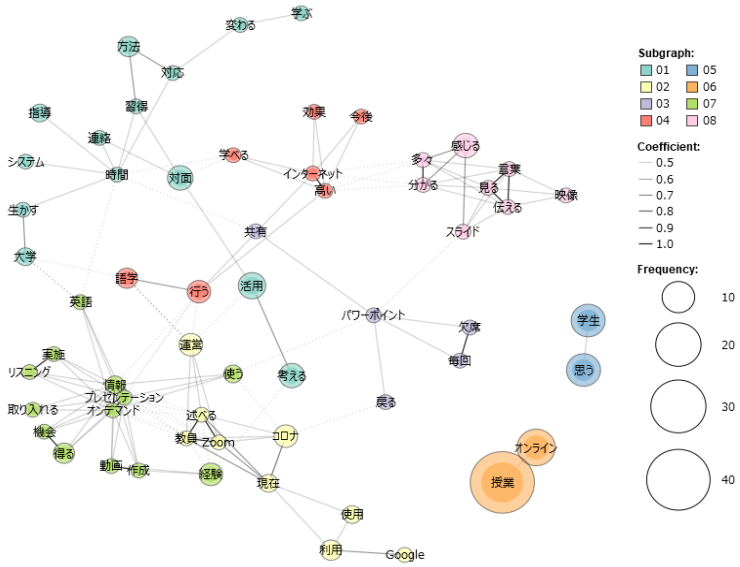
In Figure 6, online teaching as the main instructional platform is highlighted as the central experience and application. Its dominant frequency indicates that online classes became a major structural foundation for integrating other instructional skills (Figure 6).

One of the major clusters features practical digital literacy and the adaptation of educational tools. Skills such as using platforms like Zoom and online applications offered by Google were not only practiced but became part of the instructional routine. This also implies student participation, which contributed to improvements in overall computer literacy and multimodal presentation skills.

An isolated but important pair, 学生 (student) and 思う (think), seems to reflect meta-cognition and student-centered thinking about the educational environment. Another cluster, which includes 学ぶ (learn) and 効果 (efficiency), indicates perceived benefits of online formats, particularly in terms of efficiency and learning outcomes. Visual aids such as スライド (slide) played a crucial role in helping students feel, see, and understand content, pointing to effective learning enhancement. The term 欠席 (absence) seems to reflect

Figure 6

Application of Experience



efforts to ensure continuity and accessibility by helping students rejoin classes through shared materials such as visual resources. Practical applications of communicative competence through speaking activities and real-world practice are emphasized by the terms 実践 (practice), 話す (speak), 行う (do), and 活用 (utilize). It is important to note the values of visual materials such as PowerPoint slides, which links multiple clusters and serve as a bridge between emotional expression and functional continuity, especially when compensating for absences. The term 対面 (face-to-face) remains associated with learning and feedback, suggesting that hybrid environments blended smoothly with online resources.

These results indicate that educators and students actively integrated digital skills into teaching and learning practices in the post-COVID period. Tools such as Google search, Zoom, and PowerPoint became essential for content delivery, feedback, and emotional engagement. The experience of online teaching was not merely temporary, but it has been carried forward, enhancing both logistical management and affective dimensions of learning.

3.7 Correspondence Analysis: Efforts to Improve Student Satisfaction

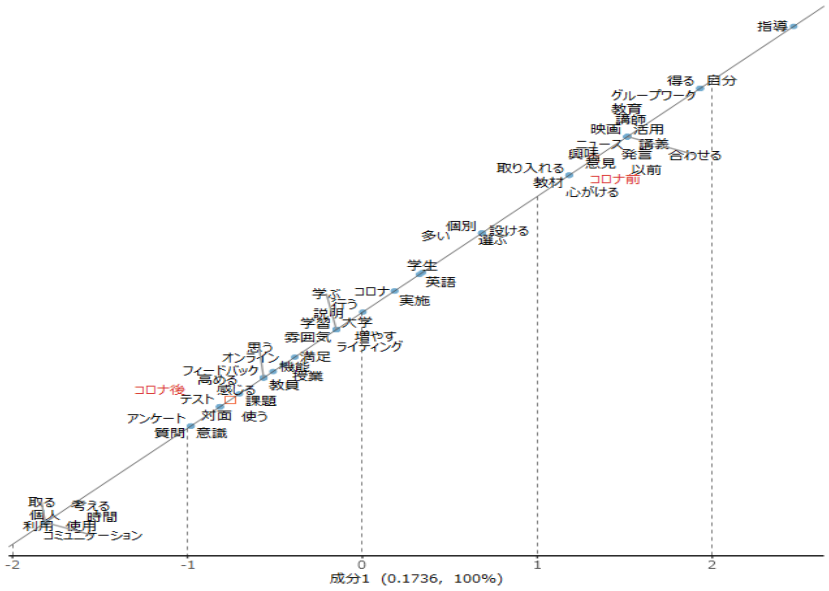
Figure 7 visually represents the correspondence relationships between key terms gathered from responses about efforts to improve student satisfaction, particularly before and after the COVID-19 pandemic. The correspondence analysis was conducted based on the frequency of co-occurring words across categorized responses, and the distances between words represent the strength of their association.

Words plotted in closer proximity indicate stronger semantic or contextual associations. One of the prominent associations in the post-COVID-19 context relates to logistical challenges, individual adaptation, and online or hybrid learning environment. Keywords such as 対面 (face-to-face), テスト (test), 使用 (use), 質問 (question), 意識 (awareness), 時間 (time), 個人 (individual), and コミュニケーション (communication) reflect a transition phase in which instructors and students were adapting to new tools, testing formats, and modes of classroom interaction.

Several central keywords span both pre- and post-pandemic contexts, including 学ぶ (learn), 実施

Figure 7

Efforts to Improve Student Satisfaction



(implementation), フィードバック (feedback), 雰囲気 (atmosphere), オンライン (online), and ライティング (writing). These terms represent unchangeable aspects of language education, such as lesson implementation and feedback processes regardless of delivery mode. In contrast, the terms such as 指導 (instruction), グループワーク (group work), 発表 (presentation), and 教材 (teaching material) are more strongly associated with pre-COVID educational experiences, reflecting a focus on active and collaborative instructional strategies such as group activities and presentations.

From right to left, a clear pedagogical shift is visible: that is, from a student-centered and interaction-focused mode before the COVID-19 pandemic to a logistical and technologically adaptive orientation afterward. The middle cluster includes terms such as feedback, students, and learning, which remain consistent across both periods and are central to the ongoing pursuit of student satisfaction. On the far right, the role and effort of teacher engagement and emotional consideration are implied by 心がける (mindful) and 指導 (instruction). This may suggest a lasting emphasis on self-awareness, guidance, and individualized support as key components of sustainable teaching practices.

4 Discussion

4.1 Pedagogical Concerns and Emotional Burden

The co-occurrence network analysis of open-ended responses revealed that part-time instructors often face both emotional and practical challenges in their teaching. The words such as pain, difficult, and task frequently appeared together, forming an important cluster that seems to reflect the mental stress associated with classroom diversity, varying student proficiency levels, unclear syllabi, and limited institutional support. Instructors noted that differences in student backgrounds and vague expectations contributed to ongoing stress and uncertainty. Another cluster included words like English, textbook, and responsibility. This group highlighted issues related to curriculum design and the limited decision-making power given to part-time instructors.

4.2 Adaptation and Coping Strategies

In contrast, clusters containing words such as method, consideration, and creativity illustrated how teachers actively adapted to different classroom situations. Participants reported using online tools such as PowerPoint slides, videos, and Google Forms, as well as hybrid teaching methods, to keep students motivated during and after the pandemic. Some instructors also mentioned providing individual support for students who missed classes or had lower English proficiency. These responses show that most of the teachers made conscious and flexible efforts to ensure the smooth continuation of their lessons.

4.3 Toward Sustainability

Another network focused on the concept of sustainability. The words such as support, innovation, and mental and physical well-being appeared in the same cluster, suggesting that teachers view sustainable education not only as a matter of teaching skills but also as dependent on manageable workloads and institutional understanding. Several instructors mentioned that heavy teaching schedules caused stress, emphasizing that continued employment and stable curricula are essential to protect both the quality of education and the teachers' well-being.

4.4 Correspondence Analysis: Shifting Student Satisfaction

The correspondence analysis comparing pre- and post-COVID-19 responses revealed clear changes in how teachers discussed student satisfaction. Before the pandemic, satisfaction was closely linked to active learning, group work, and presentation-based activities. After the pandemic, teachers used more practical and management-related terms such as assignment clarity, feedback, and online access. This shift may suggest that student satisfaction has become increasingly connected to clear communication, transparent systems, and consistent feedback rather than classroom interaction alone.

4.5 Application of Experience and Transferable Skills

Finally, the analysis of application of experience showed that teachers have continued to use digital skills developed during the pandemic in face-to-face and hybrid classes. Tools such as PowerPoint, online submission systems, and video editors are now integrated into regular teaching practices. Many instructors reported that these digital tools helped them maintain lesson continuity and clarity, particularly for students who were absent. Across all analyses, a consistent pattern emerged: that is, part-time English instructors play a vital role in English education at Japanese universities, but they often work under difficult conditions that affect both their teaching and well-being. Nevertheless, even under these constraints, most teachers demonstrated strong creativity and adaptability, especially after COVID-19. Their comments clearly show the urgent need for a more supportive and sustainable system for part-time faculty in higher education.

6 Conclusions

This study examined the experiences of part-time English instructors in Japanese higher education. The findings showed that these instructors face multiple challenges and difficulties, including job insecurity, emotional stress, and limited teaching resources. At the same time, they have developed innovative ways to address these problems, using online platforms, slides, and videos to make learning more accessible and engaging. They also strive to strengthen connections with students and enhance satisfaction through feedback and reflection.

Sustainable English education requires universities to recognize and value the contributions of part-time faculty. These findings suggest that part-time teachers can share their creativity and ideas more freely when they feel respected and supported in their teaching and working environment. This, in return, enhances not only their own teaching but also the overall quality of university English education. Ensuring fair working conditions, equitable access to teaching resources, and greater opportunities for professional development will benefit teachers, students, and institutions alike. Future research should include students' perspectives and compare practices across different universities to understand how institutional policies influence both teaching and learning. By continuing to listen to the voices of part-time instructors, a more inclusive and sustainable environment for English education in Japan can be built to further strengthen English language education.

References

Kanbayashi, Y. (2021). *Senmon shokugyō toshiteno hijōkin kōshi: Shokugyō toshiteno jizoku kanōsei o kangaeru* [Part-time lecturers as professionals: Considering sustainability as a profession]. Gakubunsha.

Kato, T. (2023). *Daigaku kyōin koyo jōkyō no bunseki to kadai* [An analysis of university faculty employment issues]. *Journal of Higher Education Studies*, 32(1), 15–28.

Kimura, S. (2016). *J-rec-in no kōkoku dēta kara miru daigaku eigo kyōiku no dōkō* [Trends in university English education from JREC-IN job postings]. *Language and Education*, 23(2), 45–59.

Higuchi, K. (2016). KH Coder 3 reference manual. Retrieved from http://khcoder.net/en/

MEXT. (2020). *Higher education during and after COVID-19: Interim report*. https://www.mext.go.jp/

MEXT. (2022). *On enhancing faculty development for non-regular university instructors*. https://www.mext.go.jp/

Taguchi, K. (2022). *FD program design for part-time faculty in the digital era*. *Journal of Educational Development and Practice*, 14(3), 55–68.

Yamaguchi, H. (2021). *Kōkō eigo kyōiku to daigaku eigo kyōiku no renkei* [Connecting high school and university English education]. *The English Teachers' Review*, 69(4), 22–30.

Appendix A

A Questionnaire Items (English and Japanese)

1. Concerns and Difficulties（不安や悩み）

What kinds of anxieties, concerns, or difficulties do you experience when teaching English as a part-time instructor?

非常勤講師として英語の授業を実施するうえで、どのような点に不安や悩みを感じておられますか。

2. Coping Strategies（対応）

How do you deal with the concerns and difficulties mentioned in (1)?

1. の不安や悩みに対してどのように対応されていますか。

3. Sustainability in Education（教育の持続性）

What efforts do you make to ensure the sustainability of education as a part-time instructor?

非常勤講師として「教育の持続性」を担保するためにどのような工夫をしていますか。

4. Improving Student Satisfaction (Pre-COVID-19)（コロナ前の工夫）

Before the COVID-19 pandemic, what strategies did you use to improve student satisfaction in your classes as a part-time instructor?

コロナ禍前は、非常勤講師として「学生の満足度」を高めるためにどのような工夫をしていましたか。

5. Improving Student Satisfaction (Post-COVID-19)（コロナ後の工夫）

Since the COVID-19 pandemic, what strategies have you used to improve student satisfaction in your classes as a part-time instructor?

コロナ禍以降現在まで、非常勤講師として「学生の満足度」を高めるためにどのような工夫をしていますか。

6. Application of Online Teaching Experience（経験の活用）

How do you think your experience with online teaching during the COVID-19 pandemic can be applied to future class management?

コロナ禍でのオンライン授業の経験が、今後の授業運営にどのように活かされると思いますか。

PaL Pan-Pacific Association of Applied Linguistics

Editorial Office

PAAL Japan Office

email: office@paaljapan.org

Selected Papers of the 29th Conference of Pan-Pacific Association of Applied Linguistics

Published by The Pan-Pacific Association of Applied Linguistics (PAAL)

First published on March 31, 2026

ISSN 2760-0904