

ISSN 1345-4854

CELE JOURNAL

NUMBER THIRTY-THREE



CENTER FOR ENGLISH LANGUAGE EDUCATION
ASIA UNIVERSITY 2025

CELE JOURNAL

NUMBER THIRTY-THREE

Editor-in-Chief

Sharon Sakuda

Editors

Tim Andrewartha

Alexis Franks

Allan Goodwin

Mark Walters

Administration

Yuko Ito, CELE Director

Mikio Brooks, CELE Deputy Director

Joshua Trued, CELE Vice-Director

A PUBLICATION OF THE CENTER FOR ENGLISH LANGUAGE
EDUCATION
ASIA UNIVERSITY
2025

COPYRIGHT © 2025 CELE, (1) The copyright of any article in this Journal belongs to its respective author or authors. (2) The copyright owners have granted Asia University and the Center for English Language Education the rights to publish and re-publish their articles in book form, as well as to digitize and post their articles online.

PRINTED IN JAPAN

CONTRIBUTOR INFORMATION

Mission Statement

The *CELE Journal* is an Asia University in-house publication which provides a forum for Visiting Faculty Members, former Visiting Faculty Members, English Department Faculty members, and AUAP consortium school scholars. The Journal seeks to present articles related to English language learning in any of the following areas: applied linguistics, second language acquisition, phonetics, psycholinguistics, teaching methodology, materials development, sociolinguistics, and global issues in language education. The *CELE Journal* is published annually in paperback form, as well as uploaded to the Internet.

Submission Guidelines

In order to assure submitting authors fair and objective assessment of their manuscripts, the Journal employs a blind reader system. Appointed readers include University English instructors from Asia University and other universities.

The *CELE Journal* accepts submissions from current and former CELE Visiting Faculty Members, English Department Faculty members, and AUAP consortium school scholars.

Submit your manuscript via e-mail to the 2025-26 CELE Vice-Director, Michael Watts: watts_michael@asia-u.ac.jp.

Submissions are generally accepted until the end of September annually. Authors may direct questions, request detailed guidelines and research subject consent forms by contacting the CELE Editorial Committee.

Requirements

All research subjects are required to fill out consent forms.

All authors are required to work with the editorial staff. The editors reserve the right to edit papers to enhance the clarity or style of the manuscript, as well as to ask for revisions to the article.

Format

All articles should adhere to the most recent APA style format. The article should be submitted as an e-mail attachment in Microsoft Word.

Contributor Copies

Ten off-prints of the individual article and one complete copy of the *CELE Journal* will be sent to the author upon publication; for multiple authors, the first named person will be sent the copies.

Online Versions

Papers printed in this volume of the *CELE Journal* shall be digitized and made open to the public through the Asia University Academic Repository on the Internet. Copyright holders shall be considered to have given permission to have their papers made open to the public in the aforesaid way. Provided that a copyright holder expresses his/her intention not to make his/her paper open to the public in the aforesaid way, it shall not be open to the public in the aforesaid way.

本紀要に掲載された論文は電子化され、亜細亜大学学術リポジトリでインターネットを通じて広く公開するものとする。著作権者は当該記載を確認のうえ電子化および学術リポジトリでの公開を許諾したものと見なす。ただし、著作権者から上述の公開を望まない旨の申し出があった場合は、当該著作権者の論文については上述の公開を行わない。

**CELE
JOURNAL
NUMBER THIRTY-THREE**

- 1 A Book Review: *Speaking by Speaking: Skills for Social Competence***
Alexis Franks
- 13 Grouping English Communication Classes by Level: Reforming Systems**
Post-COVID Pandemic Restrictions
Gordon A. Allan
- 28 Peer Feedback in Japanese EFL University Classes: Perspectives from Native**
English Instructors
Daniel Anderson
- 43 Using Gamification to Incentivise Students in Mandatory TOEIC Classes**
Denver Beirne
- 66 Effects of COVID-19 Learning Loss on Freshman English Test Scores**
Mark Goodhew
- 78 Japanese University Students' Perspective on NEST & N-NEST**
Armando Moreno & Joshua Trued
- 101 A Readability Based Comparative Study of Four Corners 1 & TOEIC Skills 1**
Joshua Trued

A Book Review: *Speaking by Speaking: Skills for Social Competence*

By David W. Dugas and Ronald T. DesRosiers

Compass Publishing, 2010, 108 pp., ISBN: 978-1-59966-571-9

Reviewed by Alexis Franks, Asia University

Abstract

This paper is a review of *Speaking by Speaking: Skills for Social Competence*. Written by David W. Dugas and Ronald T. DesRosiers and published commercially by Compass Publishing, this textbook was designed for university students who are studying English conversation at the intermediate level in a classroom setting. First, the paper presents a detailed outline of the textbook, providing information regarding each unit's theme, structure, and staging. Next, it considers the strengths and weaknesses of *Speaking by Speaking* within the context of the authors' clearly stated goals, before finally concluding that the textbook, with some caveats, is recommended for intellectually curious students who wish to improve their English communication skills.

Introduction

Speaking by Speaking: Skills for Social Competence, which was released by Compass Publishing in 2010, retails for ¥2640 in Japan. It is co-authored by David W. Dugas and Ronald T. DesRosiers, both of whom have extensive experience teaching English as an Additional Language in non-English-speaking countries. The textbook is meant to be a comprehensive and challenging guide through one semester of an intermediate English course at a university. The authors make this intention clear in the *How to Use This Textbook* segment of the book: “Though learning conversation is the primary focus of this book,” they assert, “we have made an effort to raise the level of thinking above that required by most English training books [because our] premise is that university students are intelligent, curious, and concerned young people in spite of their limited abilities to converse in English” (Dugas & DesRosiers, 2010, p. 6). In this reviewer’s opinion, *Speaking by Speaking*, despite a few shortcomings, serves as a useful resource for learners studying English as part of a post-secondary curriculum.

Overview of the Textbook

Speaking by Speaking is 108 pages long (not including a 16-page Transcripts & Answer Key perforated pull-out). The textbook has 12 units, each with eight pages evenly divided into two parts (Part A and Part B) that together comprise three hours’ worth of lesson material. Unit themes are fairly standard: *Socializing* is the theme of Unit 1, for example, and *Entertainment, Travel*, and *Health* are the themes of Units 7, 9, and 11, respectively. Topics for the parts of the units are slightly more varied, ranging from situational concepts (*Shopping at Stores*, Unit 3) to notional-functional concepts (*Reaching Agreement*, Unit 4) and topic-based concepts (*Bank Cards*, Unit 10). Each unit contains the following eight sections:

Part A

Section I. Warm-ups

This section introduces the unit’s overall theme with a title and a picture, and also announces the topic of Part A, which is directly related to the unit’s theme. Additionally, it familiarizes students with expressions that may be used to process, discuss, and personalize the topic. Section I includes three exercises:

- A. Brainstorming shows a graphic that looks like a blank piece of lined notebook paper. Students are instructed to activate their already existing knowledge (schema) by writing into the graphic as many words or expressions as they can that relate to the theme of the unit.
- B. Listening requires students to listen to a brief synopsis of the topic of Part A. They either simply listen to the audio (the textbook comes bundled with an MP3 CD) or listen while reading along with the transcript.
- C. Useful Expressions presents key expressions that will be used throughout Part A. Students are instructed to read the expressions (in blue) and match them to question or answer prompts according to how they are used in conversation. (For example, in Unit 3, the expression *How will you pay for that?* matches the answer prompt *I will use my Visa card.*)

Section II. Listening & Speaking

This section asks students to think more deeply about the topic of Part A by drawing their attention to how expressions related to the topic are used in conversation. Section II includes three exercises:

- A. Model Dialog presents the scenario and full text of a conversation between two speakers, with key expressions that were introduced in the previous section highlighted in blue. Students are instructed to listen and follow along with the written text as the teacher reads the dialog aloud.
- B. Group Work asks students to read aloud the model dialog from Exercise A. They are instructed to do this in chorus, with a different group of students taking on the role of each speaker.
- C. Guided Speaking presents the scenario and incomplete text of a conversation between two speakers (not the same speakers from the model dialog in Exercise A). Students are instructed to fill in the blanks of the conversation with any language they know, which may or may not include the key expressions that they studied earlier. Students are expected to complete this exercise in pairs.

Section III. Description

This section tasks students with reading, interpreting, and analyzing visual representations related to the topic of Part A. Following the interaction pattern established in the previous section, students are expected to complete certain activities in pairs. Section III includes three exercises:

- A. Charts/Graphs/Tables presents a visual representation (a chart, a graph, or a table) that illustrates some aspect of the topic of Part A. Students are instructed to study the representation, read the accompanying brief explanation, and reach some understanding of the information presented.
- B. Pair Work includes two mini-exercises. In the first mini-exercise, students are instructed to look at the visual representation from Exercise A again and complete sentences which describe or consolidate the information therein. In the second mini-exercise, students are expected to personalize the information in the visual representation, either by creating and describing their own chart/graph/table, or by talking about related topics with their classmates.
- C. Grammar Focus or Focus on Usage presents information about vocabulary or grammatical structures (e.g., *sequence of adjectives*, Unit 2; *expressions of degree*: kind of and sort of, Unit 7; *expression desires using I wish + past tense*, Unit 11; etc.) that may be found or used in Part A.

Section IV. Reading & Discussions

This section gives students an opportunity to practice reading comprehension by providing a longer written text that conveys facts or tells a fictional story associated with the topic of Part A. After reading the text, students answer two questions: Question 1 asks about particular information stated in the passage, and Question 2 asks about students' personal knowledge of or opinions about the information. Vocabulary prompts, prefaced with the words *Your answer may begin like ...*, are provided for each question.

Part B

Section I. Warm-ups

This section re-introduces the unit's overall theme. It also presents the topic and target expressions of Part B, which, though connected with the unit's theme, are nevertheless different from the topic and target expressions of Part A. Section I includes three exercises:

- A. Listening presents the scenario of a conversation between two or more speakers (different speakers from those in Part A). Students are instructed to simply listen to the conversation.
- B. Dictation presents a cloze exercise of the conversation from Exercise A. Students are instructed to listen to the conversation again and fill in the blanks of the exercise.
- C. Useful Expressions presents key expressions that will be used throughout Part B (and were already used in the conversation from Exercises A and B). Students are instructed to read the expressions (in blue) on the left and use them to complete the sentences on the right. (For example, in Unit 8, the expression *not at all* completes the sentence *His appearance was _____ what I expected.*)

Section II. Listening & Speaking

This section asks students to consider both the overarching theme of the unit and the topic of Part B. It also highlights the form and usage of expressions related to the theme/topic. Section II includes three exercises:

- A. Model Dialog has a similar structure and purpose to Exercise A in Section II of Part A, although the scenario, model dialog, and key expressions are unique to Part B.
- B. Group Work has a similar structure and purpose to Exercise B in Section II of Part A, although students are expected to chorally read Part B's model dialog.
- C. Guided Speaking presents a series of questions and answer prompts related to the topic of Part B. Students are instructed to use the short cues given to complete the answers to the best of their ability. They are expected to complete this exercise in pairs.

Section III. Description

This section directs students to scrutinize information from a visual representation that is linked with the topic of Part B. Again, students work in pairs to complete some of the activities. Section II includes three exercises:

- A. Charts/Graphs/Tables has a similar structure and purpose to Exercise A in Section III of Part A, although the visual representation and accompanying explanation are unique to Part B.
- B. Pair Work has a similar structure and purpose to Exercise B in Section III of Part A, although students are expected to personalize/discuss the information in Part B's visual representation.
- C. Grammar Focus or Focus on Usage presents information about vocabulary or grammatical structures (e.g., *expressions of increase and decrease*, Unit 2; *present tense in temporal and conditional clauses*, Unit 7; *participle adjectives*, Unit 11; etc.) that can be used in Part B.

Section IV. Reading & Discussions

This section has a similar structure and purpose to Section IV of Part A, although the reading passage and comprehension questions are unique to Part B.

Goals of the Textbook

The central focus of *Speaking by Speaking*, as noted in the Introduction of this review, is English conversation. This contention is supported not only by statements to that effect in *How to Use This Textbook*, but also by the fact that over half of the exercises in each unit are collaborative, with students explicitly instructed to complete the tasks together. Nevertheless, it is possible to look beyond this central *focus* to suggest that the central *goal* of the textbook is actually two-fold. The first aspect of the goal of *Speaking by Speaking* is to develop students' **communicative competence** through a sociolinguistic framework that provides both context and relatability to the conversations and exercises in the textbook. Such a format, the authors claim, "is essential to rapid [language] study" and "gives the communicative approach a reason for being" (Dugas & DesRosiers, 2010, p. 6). The second aspect of the goal of *Speaking by Speaking*, somewhat dissimilar to the first, is to enrich students' **analytical competence** by exercising their "analytic and mathematical thinking within the context of conversation" (Dugas & DesRosiers, 2010, p. 6). It is with an understanding of this dual goal that the strengths and weaknesses of the *Speaking by Speaking* textbook may be examined.

Strengths of the Textbook

Regarding Communicative Competence

Linguist Dell Hymes (1972) defines communicative competence as a language user's "knowledge of sentences not only as grammatical, but also as appropriate. He or she acquires competence as to when to speak, when not, and as to what to talk about with whom, when, where, [and] in what manner" (p. 277). Hymes (1972) argues that this competence "is integral with attitudes, values, and motivations concerning language, its features and uses, and integral with competence for, and attitudes toward, the interrelation of language with the other code of communicative conduct" (p. 277). To summarize, communicative competence concerns one's ability to use language effectively and appropriately in various social contexts.

Besides giving students a number of chances to converse with their classmates, *Speaking by Speaking* encourages the development of communicative competence primarily by providing scenarios for the conversations that appear throughout the textbook. These scenarios keep students mindful of the social functions of language by conveying "sociolinguistic details essential to guiding and limiting the proper use of English in each context" (Dugas & DesRosiers, 2010, p. 8). One example can be found in Section II of Part A in Unit 1. The scenario of the model dialog in Exercise A reads, "Scenario – Two employees from a large business firm are attending a company dinner in NYC" (Dugas & DesRosiers, 2010, p. 14). From this short sentence, students are able to ascertain several things: the relationship between the speakers, Ronaldo and Maria (they are coworkers in a large firm); the context of the conversation (small talk at a corporate event in the evening); and the location of the conversation (New York City). This knowledge enables students not only to follow the flow of the conversation (e.g., it can be assumed that the coworkers mention New York City because that is where their work event is taking place), but also to navigate its unspoken complexities. For instance, the abruptness of the dialog's first line – Ronaldo asks Maria, "*I know this is none of my business* [emphasis added], but when did you start working for the firm?" (Dugas & DesRosiers, 2010, p. 14) – may be jarring, unless one recalls from the scenario that although the speakers have something in common (they work for the same company), they are relatively unfamiliar with each other (the company is large, so it likely has many employees who are not each other's "business").

The scenario of Exercise's A model dialog in Section II of Part A in Unit 5 also illustrates how sociolinguistic information can lead to greater insight into a spoken text. It reads, "Scenario – Jose, from the Mexican countryside, is at a bar in NYC, talking with his

friend Mike” (Dugas & DesRosiers, 2010, p. 46). Details imparted here include the names of the speakers (Jose and Mike), the background of one of the speakers (Jose is from a rural part of Mexico), and the location of the conversation (a bar in New York City). These minutiae are crucial for understanding and even empathizing with the evocative metaphorical language that Jose uses to describe the homesickness he feels about being away from his homeland: “I am out of my element,” he says. “My heart is heavy for home” (Dugas & DesRosiers, 2010, p. 46).

In fact, Unit 5’s scenario-dialog combination exemplifies the comprehensiveness and complexity with which *Speaking by Speaking* deals with sociolinguistic factors that determine how language is used in different social settings. Specifically, it portrays a meaningful, cross-cultural interaction between Mike, a character from the United States, an Inner Circle country (where English is spoken by a majority of people as their first language) and Jose, a character from Mexico, an Expanding Circle country (where English is recognized as a *lingua franca* and is widely studied as an additional language). The dialog even goes so far as to have Mike tease Jose about his yearning for Mexican comforts and the potential difficulty of finding them in the U.S.: “How about going to Taco Bell [an American fast food chain] for some authentic Mexican food?,” he asks. “Ha, ha, ha, just joking” (Dugas & DesRosiers, 2010, p. 46). As Sercu (2002) points out, studying an exchange like this, which “unveils the fissures, inequalities, disagreements, cross-cutting influences, as well as the agreements and elements of stability that exist in and around all cultures” (p. 68), is a vital component in developing communicative competence.

Regarding Analytical Competence

According to Bloom’s Revised Taxonomy (Anderson & Krathwohl, 2001), students achieve analytical competence when they are able to deal with complex information in the following ways: first, by breaking down the information into its component parts; second, by identifying the relationships of the parts to each other and to the whole; and third, by reaching conclusions or making decisions about the information. Each unit of *Speaking by Speaking* includes multiple opportunities for students to attempt these tasks. Such opportunities are chiefly concentrated in Section III of both Part A and Part B, where complex information – in the form of a visual representation – is presented for comprehension and examination.

Part B in Unit 1, showcasing a timetable that outlines the weekly schedule of Barbara, a university student majoring in Business, provides a functional example of how analytical

skills are integrated into the textbook's Section III exercises. In Exercise A, students are encouraged to "learn how busy college students are" (Dugas & DesRosiers, 2010, p. 19) by studying the timetable carefully. Then, in the first mini-exercise in Exercise B, they are asked to complete three sentences (a. – c.) based on the information. Each sentence consolidates various elements of the timetable's content, thus requiring students to deduce how the different components of the timetable relate to one another and to Barbara's entire schedule. For example, in order to successfully complete Sentence a. (*Barbara always _____ at 7:00 a.m. weekdays.*), students must grapple with the fairly disparate concepts of grammar (adverbs of frequency such as *always*), lexis (the difference between *weekdays* and *weekends*), and mathematics (the conversion of the table's 24-hour expression of time to the sentence's 12-hour expression of time).

The second mini-exercise in Exercise B requires further analysis of the timetable. Following the instructions to "compare your schedule with Barbara's, and share your results with a partner" (Dugas & DesRosiers, 2010, p. 19), students must not only address the complicated question of how their own schedules relate to Barbara's, but they must also come to some conclusion(s) about this relationship (e.g., whether they wake up earlier than Barbara; whether they eat dinner at the same time as Barbara; whether they are busier than Barbara overall; etc.) and determine how to express their conclusion(s) in English.

Weaknesses of the Textbook

Regarding Communicative Competence

Speaking by Speaking cultivates students' communicative competence in several ways. Most meaningfully, it presents scenarios and conversations that shine a spotlight on sociolinguistic and intercultural factors which contextualize and complicate the use of language. Despite this, the textbook detracts somewhat from its communicative goal by curtailing activities that would incentivize students to work together to achieve some objective. To illustrate, although all of the units in *Speaking by Speaking* include exercises that were ostensibly designed for students to collaborate, many of them actually require very little coordinated effort to fulfill the tasks at hand. Take as a model Exercise C in Section II of Part A in Unit 5. This exercise presents an incomplete conversation between two speakers – Frank and a character identified only as Stranger – who are meeting at a bar for the first time. Students are expected to read the gapped text of the conversation, fill in the blanks with appropriate English expressions, and perform the full conversation while switching roles with

a partner. However, in defiance of the directive to act as a team, students could easily do these tasks on their own: they could simply complete the conversation with their own ideas and then stop working altogether, or, if inclined to continue, they could just read aloud their completed text to their partner (rather than engaging in a role-play of the conversation).

As they stand, activities such as the ones described above, which are peppered throughout *Speaking by Speaking*, do not adequately promote pair or group work; it may even be argued that they “merely [put] the words ‘In groups’ or ‘In pairs’ in front of ... individual activities, without making any changes to encourage learners to cooperate with each other” (Jacobs & Ball, 1996, p. 1). Yet, certain amendments to these activities could provide incentives for students to cooperate, and thus more assuredly enhance students’ communicative skills. For instance, the first instructions in Exercise C above are: “Complete the dialog and practice it, changing roles with a partner” (Dugas & DesRosiers, 2010, p. 46). If these instructions were replaced with something like *Work in pairs. Take turns to complete each line of the dialog and justify your ideas to your partner.*, students would be better able to recognize that it is only the combination of both partners’ contributions (and their ability to communicate these effectively) that leads to successful completion of the activities.

Regarding Analytical Competence

Although *Speaking by Speaking* refines students’ analytic skills to a certain extent, namely by encouraging them “to learn to identify and use specific information given in the graphics [visual representations]” (Dugas & DesRosiers, 2010, p. 9) in Section III of Part A and Part B in each unit, the textbook does not do enough to ensure that conclusions or decisions are made about the information that students consume. More explicitly, out of the book’s 24 sets of instructions for Exercise B in Section III, no fewer than 17 include the somewhat simplistic verbs *interview*, *share*, and *report*, in spite of the fact that the ‘Analyze’ band of Bloom’s Revised Taxonomy (Anderson & Krathwohl, 2001) contains a surfeit of additional verbs (e.g., *discuss*, *organize*, *distinguish*, *question*, *infer*, etc.) that could provide foundations for a more sophisticated analysis of content.

Again, certain alterations of the textbook’s exercises could guarantee that *Speaking by Speaking* more fully accomplishes its aims. Regarding analytical aims, the incorporation of extra Section III tasks could enable students to interact with the textbook’s visual representations in a more nuanced and logical fashion. For instance, in Section III of Part A in Unit 1, a pie chart depicts how university students in the U.S. spend their time over the

course of 24 hours. Students study the pie chart in Exercise A and complete three descriptive sentences about the information in the first mini-exercise in Exercise B. Afterwards, they complete two tasks in the second mini-exercise in Exercise B. The instructions for the tasks are: “2. Show how you use your time on an average weekday by filling in the following table. Then create your own pie chart. 3. Compare your typical day with a partner’s” (Dugas & DesRosiers, 2010, p. 15). As scrutiny of the information in the pie chart stops at making comparisons (according to these instructions), the addition of a third task at this point might successfully facilitate deeper analysis. Utilizing some of the taxonomic adjectives listed above, instructions for this supplemental task could be something like 1. *Distinguish between the typical day of a university student from the U.S. and the typical day of a university student from your culture: Which student is busier?* 2. *Question your conclusion: Can you think of any culture-based reasons why one student is busier than the other?* 3. *Discuss your ideas, along with any supporting evidence, with your partner.* Bringing students’ analytical capabilities to bear in this way would, according to Khodos and Hunt (2022), “help students view language objectively and scientifically [and] provide a critical ... reflection upon actual language” (p. 154).

Conclusion

Speaking by Speaking: Skills for Social Competence is largely successful in fostering students’ communicative and analytical abilities. To that end, the textbook contains numerous texts and exercises that emphasize sociolinguistic, intercultural, and discursive features of language, and it introduces a variety of thought-provoking visual representations and activities that motivate students to respond fairly logically to complex information. Nonetheless, *Speaking by Speaking* does have some slight limitations (i.e., omissions) that prevent a complete realization of its authors’ twofold goal. In spite of its imperfections, however, the *Speaking by Speaking* textbook capably “stimulates [students’] overall intellectual growth, as well as their English conversation skills” (Dugas & DesRosiers, 2010, p. 6).

References

- Anderson, L. W. & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching and assessing: A revision of Bloom's taxonomy of education objectives (complete edition)*. Longman.
- Dugas, D. W. & DesRosiers, R. T. (2010). *Speaking by speaking: Skills for social competence*. Compass Publishing.
- Hymes, D. (1972). On communicative competence. In J. Pride & J. Holmes (Eds.), *Sociolinguistics* (pp. 269–285). Penguin Books.
- Jacobs, G. M. & Ball, J. (1996). An investigation of the structure of group activities in ELT coursebooks. *ELT Journal*, 50, 99–107.
- Khodos, I. & Hunt, J. (2022). Developing students' linguistic and analytical skills: The use of anchored instruction in an introductory sociolinguistics course. *International Journal of Teaching and Learning in Higher Education*, 34(1), 153–160.
- Sercu, L. (2002). Autonomous learning and the acquisition of intercultural communicative competence: Some implications for course development. *Language, Culture & Curriculum*, 15, 61–74.

Grouping English Communication Classes by Level: Reforming Systems Post-COVID Pandemic Restrictions

Gordon A. Allan, Asia University

Abstract

Asia University's Centre for English Language Education (CELE) offers a suite of elective English Communication (Comm) courses. In 2020, a move from fifteen to thirteen-week semesters coincided with the start of the COVID-19 pandemic. As a result, the first year of thirteen-week semesters was taught entirely online, causing significant disruption for students, teachers, and administrators alike. Under the old fifteen-week system, level checking for Comm classes took place in the first week of each semester; a "shopping" week in which students could try out and sign up for classes. The shopping week disappeared in the new thirteen-week system and one consequence of the disruption caused by COVID was that no alternative level-checking system was established for thirteen-week Comm courses. Students were instead trusted to self-assess. This was compounded by an apparent failure to effectively communicate information about Comm classes to students, resulting in some students applying for classes above or below their English level. Over time, Comm classes effectively became mixed-level in practice, causing difficulties for both teachers and students. Beginning in 2023, CELE's Curriculum Development Committee (CDC) worked with CELE management and the Registration division to attempt to rectify the situation, first by conveying more specific information about Comm courses to students and subsequently by implementing a level-checking system based on students' TOEIC IP test scores. A detailed, comprehensive process was negotiated, trialled, and then put into action for the first semester of 2024-25. Tight deadlines were met for both rounds of applications, and the resultant allocations had a much narrower range of TOEIC scores than would have been the case under the previous system. Feedback from teachers suggests that this has improved the quality of Comm classes overall.

Introduction

The present paper is concerned with the “English Communication” suite of communicative English language classes at Asia University, a private university in Tokyo with approximately 6,600 students. English Communication (Comm) courses are elective, once-a-week classes delivered by Asia University’s Center for English Language Education (CELE), which provides English teaching services to all faculties in the university. In academic year 2024-25, fourteen Comm courses were offered in the first semester and nineteen in the second semester. Comm courses attempt to offer an appropriate range of courses each semester to meet the needs and interests of Asia University students, and they constitute a self-contained system within the broader range of courses taught by CELE teachers.

O’Sullivan’s (2020) model of a comprehensive learning system (CLS) identifies three elements that need to be in harmony for an educational system to succeed: curriculum, assessment, and delivery. For Comm classes at Asia University, the specifics of curriculum and assessment for each class are decided by teachers, who are free to use a textbook of their choice (provided it is not already being used for non-elective courses elsewhere in the university) or to create their own materials. They are also free to decide the specific aims of the Comm courses that they teach, provided those aims fall within the broad category of English communication. The curriculum and assessment elements of the system are therefore very loosely defined at the university level. There is, however, a structure that divides Comm courses into four levels approximately equivalent to elementary (I & II), pre-intermediate (III & IV), intermediate (V & VI), and upper intermediate and above (VII & VIII) (the Roman numerals denote semesters – odd numbers for the first semester and even for the second).

O’Sullivan (2020) notes that, “Traditionally, the delivery system has been seen as a process by which the formal curriculum is operationalised in specific learning contexts or domains” (p.10), and breaks this down into three categories: physical environment, school staff, and learning materials. Although not mentioned in the model, in addition to staff, the other group of people that influence learners’ experience of a course is other students in the class, especially in a communicative language class. This paper therefore posits that administrative procedures may also play an important role in the delivery of language courses by determining how places are allocated to prospective students.

The paper begins by outlining how a change in the processes for allocating places on Comm courses had a knock-on effect on the curriculum. It then describes measures that were

taken to implement a new system for assigning places and teachers' perceptions of the extent to which these addressed some of the issues. Finally, there is a brief discussion of underlying issues, and future actions are suggested to build upon the reforms already undertaken.

Background

In 2020, Asia University transitioned from fifteen-week semesters of 90-minute lessons to thirteen-week semesters of 105-minute lessons, the increase in lesson length compensating for the shortening of the semester in terms of overall teaching time. Under the old fifteen-week system, the first week of each semester was a "shopping" week in which students could try out different classes and choose what they wanted to study. Teachers presented their courses to students and were able to give prospective students short assignments to gauge their suitability for the course. This allowed language teachers to select students appropriate for the level of the class. One consequence of the move to a thirteen-week semester was the loss of this initial "shopping" week.

It was unfortunate that this change coincided with the onset of the COVID-19 pandemic. Like universities throughout Japan (MEXT, 2020), Asia University delayed the start of the first semester. When teaching did eventually start, it was entirely online for the whole of the 2020-21 academic year. Adapting face-to-face courses for online teaching placed a huge burden on teachers and institutions. Under these unusually trying circumstances, devising a new level-checking system was superseded by other priorities. In the absence of a custom-designed alternative, Comm courses ended up defaulting to the university standard for allocating places on elective courses, which meant that all applicants were accepted up to a maximum class size of 35, with places on over-subscribed courses allocated by lottery. Students were therefore trusted to self-assess and apply for appropriate courses without any level checking.

As pandemic restrictions were lifted in 2021-22, face-to-face teaching gradually returned to the classroom (Allan & Bryden, 2023). The lottery system for allocating places on Comm courses, however, continued, as no alternative system had ever been devised to replace the "shopping" week. From an administrative point of view, the lottery system was working well since it appeared to allocate places fairly with a minimum of fuss. However, anecdotal evidence from teachers suggested that all was not well in the classroom. Despite Comm classes being grouped by level on paper, in practice, many teachers complained of having to teach a mixture of different levels in the same class. It seemed that many students

were not reliably selecting classes at an appropriate level for themselves. Instead, factors such as their own timetable and their desire to be in the same class with their friends seemed to be influencing students' choices.

In response to this anecdotal evidence, the Curriculum Development Committee (CDC) surveyed teachers to ascertain whether there was any substance to it. Results indicated a widespread perception among teachers that Comm classes were in effect mixed level and that this was an obstacle to good teaching, particularly in higher level classes, in which some students were well below the expected level and required a lot of support from the teacher and other students just to cope with the material.

With this evidence that mixed-level classes were causing problems for both teachers and students, the CDC decided to comprehensively review CELE's Comm offer with a view to reforming it. A number of issues with Comm courses (detailed below) were identified that might have been contributing to the problems reported by teachers. The CDC then attempted to implement solutions. Following the CDC interventions, Comm teachers were surveyed, and follow-up interviews were conducted to assess teachers' perceptions of the extent to which these interventions had been successful. Surveys were sent via Google Forms to all thirteen teachers who taught Comm classes in the first semester of 2024-25. Nine of the thirteen responded, seven of whom indicated that they were willing to be interviewed. Follow-up semi-structured interviews were conducted individually on Zoom with all seven. To preserve anonymity, survey respondents and interviewees have been allocated pseudonyms for reporting purposes: Aaron, Bob, Colin, Damien, Emma, Freddie, Gloria, Harry, and Isla.

Comm Course Levels

The range of Comm classes offered at Asia University in 2024-25 is summarised in Table 1:

Table 1
Comm Courses at Asia University in Academic Year 2024-25

Asia University level	Description on paper	CEFR Equivalent (in theory)	TOEIC Equivalent (in theory)	Classes in Semester 1	Classes in Semester 2
Comm I & II	Elementary	A1	120-220	8	8

Comm III & IV	Pre-intermediate	A2	225-545	3	8
Comm V & VI	Intermediate	B1	550-780	2	2
Comm VII & VIII	Upper-intermediate - Advanced	B2 – C1	785-940 945-990	1	1

Eight of the fourteen classes offered in semester 1 were nominally elementary, which is commonly understood to equate to A1 on the CEFR (e.g. British Council, n.d.) and is approximately equivalent to TOEIC 120-220 (Tannenbaum & Wylie, 2019). TOEIC scores are confidential, so they will be reported here only in broad terms, but analysis of TOEIC IP test results showed that the vast majority of Asia University students are above the 120-220 range. It therefore appears that there are too many Comm I & II courses scheduled and too few courses above that level to meet the needs of the Asia University student population.

Although this seems an obvious problem, it is complicated by the levels of classes in practice not necessarily matching the levels on paper (see below). The CDC therefore decided to deal with other issues first. A proposal for re-organising Comm course levels was drawn up in November of 2023, but it has yet to be implemented and may need to be revised depending on the outcome of other interventions (see discussion).

Comm Course Descriptions

When Asia University students are deciding which courses to apply for, they are asked to refer to course descriptions on Adai Portal, a university web resource. Although there is a facility to upload separate descriptions for each course, in academic year 2022-23 and in years prior, all Comm courses had the same generic description regardless of level. This meant that prospective students had no way of knowing exactly what the course content might be prior to signing up. In addition, the absence of information about expected English level and textbooks used meant that the Roman numeral system was the only way for students to gauge the level of the course. It is not known how well students understood the relationship between the Roman numeral system and the level of English required to participate, but it seems likely that this was opaque to many prospective students.

CDC's first intervention was, therefore, to ask teachers to write course descriptions using the Adai Portal format for their Comm classes in the 2023-24 academic year. These descriptions then replaced the generic descriptions on Adai Portal. Even though there was still no level checking at this point, Aaron and Isla both reported that their classes were more tightly grouped than in 2022-23. Aaron commented that students' levels had been "spread more evenly across the entire spectrum" prior to 2023.

The apparent success of this relatively simple intervention suggests that one of the factors contributing to mixed level classes in previous years may simply have been a lack of information for students about class levels or ineffective communication thereof.

No Level-Checking for Comm Courses

As noted above, there has been no level-checking system in place for Comm courses since 2020 and teacher feedback indicated that, over time, many classes had become mixed-level in practice under this arrangement. Some students may have been unaware of the level of the class(es) they were applying for or may have prioritised other factors such as their schedule or wanting to study with their friends.

As Comm classes became mixed-level, teachers were forced to adopt coping strategies, one of which was to aim classes at a perceived "middle" level. This appears to have resulted in the levels of some courses drifting over time, with the distinction between levels being eroded, particularly for Comm III & IV and Comm V & VI. Evidence for this can be seen in the textbooks chosen. For example, *Unlock 2*, rated by the publisher as suitable for A2 on the CEFR (Dimond-Bayir et. al., 2019) is used to teach Comm III & IV (A2) and Comm V (B1) classes. Similarly, *Keynote 2*, rated as A2 – low B1 (Bohlke, 2017), has been used to teach Comm II (A1), Comm III & IV (A2), and Comm V (B1).

Since the levels of Comm classes on paper did not appear to be a good match for the levels of Asia University students in practice, there may have been good reasons for teachers to make these choices. So, rather than imposing a top-down re-organisation of levels, CDC decided to use the levels that the courses were actually being taught at as the starting point for a bottom-up reorganisation of the levelling system. This risked some inconsistency to begin with but had the advantages of being minimally disruptive and based on the reality of what was happening in classrooms. Any attempt to re-establish a level system for Comm courses was necessarily going to involve having some control of the levels of the students in a class, so CDC embarked on a project to devise and implement a level-checking system.

Implementing a New Level-Checking System for Comm Courses

Each semester, several hundred students apply to take Comm classes. There are two rounds of applications separated by a one-week interval. For each round, the turnaround time between applications being received and allocation data being returned to the Registration division is approximately 24 hours, sometimes less (typically from early or mid-afternoon until noon the following day). As a result, individually level-checking each applicant is logistically impossible, so another way of gauging students' levels is required. This is one reason why a lottery system had been preferred, because it was relatively simple and could be completed quickly. In addition, a lottery system was seen as being fair. Any new system had to be similarly swift and completed in such a way that decisions could be justified to any student who sought to challenge the outcome.

Since Asia University offers all students the opportunity to take the TOEIC IP test several times throughout their time at university, using TOEIC scores was an obvious solution. The TOEIC IP test only assesses the receptive skills of reading and listening, so it is not an ideal fit with Comm classes, which often involve a speaking focus. Since the alternative was a lottery system, however, any indication of English ability was considered better than none.

Implementing a new system involved gaining approval within the university and working closely with the Registration division, whose responsibility it is to process applications and allocate places to students. The CDC drew up a summary for the CELE leadership team, outlining the issues with Comm courses and proposing the use of TOEIC scores as a basis for making placement decisions. This was presented for approval first to the English teaching faculty and then to the Registration division. Once approved, Comm courses were added to a list of non-lottery-allocated classes.

Devising a process for allocating places

Having secured the necessary agreement within the university, the CDC then negotiated a detailed process with the Deputy Director of CELE, which involved making provisions for several different "what if" scenarios, such as how to deal with students who don't have a TOEIC score (see below) and how to accommodate students who need Comm credits to graduate. Working through the what ifs in advance helped to save time on the day because a plan existed for every eventuality that had been discussed.

Next, a small team of teachers who had expressed an interest in the project was invited to help process the applications, and a trial run using a dummy data set was conducted. Feedback from the team was used to streamline the process before receiving the real data. The final agreed-upon process is summarised below:

1. When the application data arrives, it is a single list of all applications in an Excel spreadsheet. First, copy the data for each class to a new tab such that there is one tab for each class.
2. Next, sort the data for each class by TOEIC score.
3. Check the TOEIC score range for each class and enter it in the header of the next column on the spreadsheet. Then, mark the applicants who are within the range by colouring the cells in the TOEIC range column green and entering numbers starting from one (entering numbers allows you to easily count the number of students who have been accepted). If there are more applicants within the specified range than places available, allocate places by lottery from within this group.
4. If there are still places available on the course, mark the applicants who have TOEIC scores within +/- 50 points of the specified range by colouring the cells light green and continuing the numbering. Accept the students closest to the specified range first.
5. If any places are still available, students outside the range can be considered on the following basis: For Comm I & II classes, only students below the specified range may be considered (which prevents more proficient students from potentially dominating lower-level classes). For Comm III – VIII classes, only students above the specified range may be considered (which ensures that a range of classes and times are available for the highest-level students, who would otherwise only be able to apply for one or two classes each semester).
6. Check a list of priority students (provided by Registration) who need Comm credits for graduation and include them even if they are outside the specified range of the class(es) for which they applied.
7. Review the allocations class-by-class as a team, discussing any uncertain cases to reach a consensus on which applicants to accept.
8. Finalise the class lists by entering “O” for accepted applicants and “X” for rejected applicants in the designated column.
9. Return the finalised spreadsheet to the Registration division.

Although uptake of the TOEIC IP test for freshman students exceeds 85%, one concern was that some prospective Comm students would not have a TOEIC IP test score recorded by Asia University, either because they had declined the opportunity or because they were coming from another institution and had yet to be given the opportunity to take the test. For these students, a Google form was created that allowed them to submit evidence of other English qualifications they might have along with a statement of their reasons for wanting to take a Comm course. Conversions between different tests are problematic because different tests do not all test the same things, but some indication of level was considered better than no indication at all, so the following approximation was used based on published online conversions from a variety of sources (Cambridge University Press and Assessment, 2024; Eiken, n.d.; ETS, n.d.; Tannenbaum & Wylie, 2019):

Table 2

Approximate Conversions between Different English Language Test Results

Level	CEFR	TOEIC	Eiken	TOEFL PBT	TOEFL iBT	IELTS	TEAP	Cambridge
Elementary	A1	120-220	5, 4, 3	-	-	1.0-2.5	-	-
Pre-Int	A2	225-545	Pre-2	350	20-41	3.0-3.5	135-224	Key KET
Intermediate	B1	550-780	2, 2A	450	42-71	4.0-5.0	225-308	Preliminary PET
Upper-Int	B2	785-940	Pre-1	550	72-94	5.5-6.5	309-374	First Certificate FCE
Advanced	C1	945-990	1	600	95-120	7.0-8.0	375-400	Advanced CAE
	C2	-	-	-	-	8.5-9.0	-	Proficiency CPE

Deciding TOEIC score ranges for classes

Having agreed upon a process for using TOEIC scores to allocate places, the next challenge was to determine the appropriate range of TOEIC scores for each class. Ideally, each level would be matched to a range of TOEIC scores. As discussed above, however, the levels on paper did not necessarily match the levels in practice, and the levels themselves had become ill-defined. So, rather than imposing TOEIC score ranges on these classes, teachers were asked to specify what they thought was an appropriate range of scores for the class(es) that they taught. One advantage of this bottom-up approach was that it allowed teachers to continue using the materials that they had selected and/or developed for their classes. However, there were also considerable disadvantages. This approach required teachers to be

familiar with how TOEIC scores mapped on to the English levels they encountered in class, which, understandably, not all teachers were. It also meant that classes at the same level on paper specified different TOEIC ranges in practice. This was potentially confusing, but as an interim solution, the benefits were considered to outweigh the drawbacks. The ranges will be subject to further revision as the re-organisation progresses, however, with the eventual aim of having a single defined TOEIC range for each Comm level.

Feedback from Comm teachers

Students with lower TOEIC scores were prioritised for Comm I classes. Of the four teachers teaching Comm I, two were satisfied or moderately satisfied with the range of levels in their classes, while two were neutral. Interestingly, the two more satisfied teachers had specified TOEIC ranges with little overlap between CEFR boundaries (one clearly A1 and the other clearly A2), while the neutral teachers had specified broader ranges that crossed over from A1 into A2. This feedback may therefore indicate that it is particularly important to demarcate the A1/A2 boundary for future courses.

Bob, who taught the class that was clearly at the A1 level, spoke in the interview about the benefits of not having higher-level students in his class, commenting that if “one person’s really good at English and the other person is not very good at English, then the person who’s not good at English tends to become really shy and reluctant.” He also observed that, “If a higher-level student is with a group of this mostly lower-level students, they often tend to adjust to that because they don’t want to stand out, and so they would actually bring their level down. And I kind of feel like that’s not really benefiting anyone in that kind of scenario. But I didn’t see that within this group of lower-level students ... I didn’t see them adjusting and going lower. So really, what I felt was they were bringing up the students who were lower level because they were kind of in that same range.” These insights suggest that a tighter range of levels may help to make the lower-level students feel more comfortable and simultaneously harness the potential of the higher-level students to raise the level of the group.

Students with higher TOEIC scores were prioritised for Comm III, V, and VII classes. Survey data indicated that four of the five teachers teaching Comm III, V, or VII were satisfied or moderately satisfied with the range of levels in their classes, with the other teacher expressing moderate dissatisfaction. The dissatisfied teacher complained of a persisting range of levels in the class with no noticeable change from previous years and

mentioned that the “class size is too big as always,” a complaint that was echoed by other teachers. Emma, Dominic, and Isla, however, all felt that their classes were more tightly grouped than in previous years, with fewer lower-level students who struggled to cope. Dominic noted that, “This has made planning and pacing more manageable on my part. In fact, I was able to cover all the topics in my course outline/plan, which I was not able to do last year.”

In general, teachers reported that they felt able to deal with a few outliers provided most of the class was grouped at approximately the same level. The problems they experienced with mixed-level classes tended to occur when there was an even spread of levels through the class, or “all over the entire map,” as Aaron put it. The TOEIC scores of previous years’ applicants are not known, but this year’s applications for Comm I classes included ranges such as 135-865, 140-765, 155-750, 210-990, 185-690, and 125-830, all of whom would have been accepted under the previous system because there were 35 or fewer applicants. Similarly, the two Comm V classes had applicant TOEIC ranges of 190-755 and 140-990, all of whom would have been accepted previously. Colin summed up the problem with a broad, evenly spread range of levels: “I don’t have that much time for scaffolding the lower-level students because the higher-level students are going to get bored. And there’s a lot of people to keep active.”

Emma and Isla noted a positive effect on attendance in their classes, with Emma commenting: “Compared to last year (2023-24), this year’s group seems to have similar levels (with the exception of a few students). Last year’s group was very mixed. Attendance this year is quite a bit better, with only a few students dropping (4 out of 35).” This is to be expected if students are more accurately grouped and placed in classes appropriate to their level. If confirmed, it represents a more efficient allocation of places for over-subscribed classes.

Overall, teacher’s feedback on using TOEIC scores to level-check students for Comm classes was broadly positive. Comments such as, “Things are definitely better,” (Harry) and “Thanks for doing this because I feel like the Comms classes are going in the right direction,” (Emma) suggest that the new system is an improvement on what came before.

Discussion: Class Size

Although not specifically asked about in the survey, one issue that was mentioned by most survey respondents was class size. In semester 1, nine of the fourteen classes had 34 or 35 students in them (35 being the maximum permitted). One teacher’s class was so large that

it had to be moved to a bigger room after the first week. Because class size was mentioned so consistently in the survey, interviewees were asked what they thought the ideal maximum number of students in a Comm class should be. The almost universal consensus was that about 20-25 would be a more realistic upper limit. Freddie commented that, “If it were a larger class, it would be very challenging for me to understand the different levels of the students,” a sentiment echoed by Dominic and Emma. Emma also mentioned the difficulty of doing any activities which involve moving around the room: “When I have 35, you can’t do physical activities. You have to just sit there, which, personally, I don’t like. But I mean, what can you do?”

The 35-student limit is a university standard for elective classes, and communicative language classes are treated no differently than lectures. There are, however, significant differences between communicative language teaching and lecturing, with the former requiring more interaction and a greater variety of interaction patterns.

In the assessment community, language assessment literacy (LAL) has been the focus of much discussion and research in recent years (e.g., Coombe et.al., 2020). There is growing recognition that a variety of stakeholders need an understanding of how language assessment works and how test results can be used to make often important decisions in contexts such as employment, university admissions, and immigration, for example. It is therefore incumbent on those involved in language testing to help these stakeholder groups develop a level of language assessment literacy appropriate to their needs and responsibilities.

The issues with level checking and class size described in the present paper suggest that there may be a parallel in communicative language teaching, albeit on a smaller scale. Both issues seem to have stemmed from a lack of understanding of how communicative language classes differ from more traditional lectures. Perhaps we as language teachers need to be raising awareness of what we do and how to help colleagues in other faculties and departments understand why mixed-level classes can be a problem for both teachers and, more importantly, students and why the maximum class size might differ from that of a traditional lecture.

Conclusion

The present research has highlighted the importance of providing information about individual courses, ideally in terms that non-specialists can understand, so that students can make informed choices. There is, however, clear evidence that not all students can be relied

upon to self-assess and select classes at an appropriate level for their English ability. The absence of a level-checking system from 2020–2024 appears to have resulted in many Comm classes becoming mixed-level, which forced teachers to adopt coping strategies. As a result, courses tended toward a perceived middle level, and the distinction between levels on paper became eroded in the classroom. This indicates that a level-checking system is required to avoid classes becoming mixed-level in practice.

The most obvious problem associated with mixed-level classes is perhaps that some students may not be proficient enough to cope with the material being taught. In addition, feedback from teachers also indicates that students whose English level is above that of a class may upset group dynamics and inhibit lower-level students, for whom the course is intended. Both of these possibilities seem to be more likely to occur in classes where there is an even spread of levels throughout the class. However, if a class is more tightly grouped around one particular level, teachers found the presence of a few outliers to be less problematic. Teachers' comments suggest that higher-level outliers can be incorporated more productively into more tightly grouped lower-level classes, while lower-level outliers can be supported more effectively in a more tightly grouped higher-level class. Some teachers also report that the implementation of a level-checking system has resulted in fewer students dropping out of classes. If confirmed more widely, this would indicate a more efficient allocation of places than the lottery system had achieved.

The bottom-up approach to this reorganisation helped to capture teachers' knowledge and experience of working with their students but resulted in some inconsistency. Going forward, some consolidation may now be needed to produce an internally consistent framework. It may be desirable to provide more guidance, possibly by specifying ranges for different levels and asking teachers to choose a level for their class rather than giving them a completely free choice. This would help to keep each level distinct, which feedback indicates may be particularly important at the A1/A2 boundary. Alongside this, reform is likely to require making the naming system for Comm levels more transparent and re-naming some classes to more accurately reflect the level at which they are being taught.

The issue of class size which arose during this research is another potential focus for future reform. Feedback from teachers suggests that classes of 35 may reduce the range of activities and interaction patterns that are practical in a lesson, as well as compromise teachers' ability to monitor effectively. Teachers of English as a Foreign Language will immediately recognise these as important aspects of a communicative language lesson, but

they may be less evident to administrators and/or colleagues in other faculties. In order to make the case for smaller class sizes, it may be necessary for CELE to proactively raise awareness of the nature of the teaching it provides. In addition, the high demand for places on Comm courses in the first semester may be a barrier to reform. Any reduction in class size will need to be accompanied by the provision of additional classes if capacity is to be maintained or expanded.

Revising the Comm course suite presented a complex set of problems. The common thread running through them seems to have been that default standards for academic lectures are not necessarily optimal for communicative language classes. From an administrators' point of view, pedagogical issues such as those outlined above may not be apparent unless complaints are received from students. Concerns about fairness and meeting deadlines are more immediate and so more likely to be prioritised. Ultimately, however, administrators are concerned with the quality of education provided to students and were receptive to the case that was made for reform. By negotiating a process that met administrators' requirements, the CDC was able to address pedagogical concerns to the benefit of teachers and students alike. It is hoped that this study is of help to anyone considering similar reforms in future.

Acknowledgements

My thanks to all the CELE teachers who responded to the survey and gave up their time to be interviewed for this research. I would also like to thank my CELE and CDC colleagues Mikio Brooks, Joshua Trued, Michael Watts and Mark Goodhew for their support and advice throughout the project.

References

- Allan, G., & Bryden, S. (2023). Blended synchronous learning: Integrating remote and face-to-face learners. In P. Ferguson, B. Lacy, & R. Derrah (Eds.), *Learning from Students, Educating Teachers—Research and Practice*. JALT.
<https://doi.org/10.37546/JALTPCP2022-25>
- Bohlke, D. (2017). *Keynote 2*. Cengage.
- British Council. (n.d.). *Understand your English level*. Retrieved December 11, 2024, from <https://learnenglish.britishcouncil.org/english-levels/understand-your-english-level>
- Cambridge University Press and Assessment. (2024). *The Cambridge English scale*.
<https://www.cambridgeenglish.org/exams-and-tests/cambridge-english-scale/>
- Coombe, C., Vafadar, H., & Mohebbi, H. (2020). Language assessment literacy: What do we need to learn, unlearn, and relearn? *Lang Test Asia*, 10(3).
<https://doi.org/10.1186/s40468-020-00101-6>
- Dimond-Bayir, S., Russell, K., Sowton, C., Farmer, J., & Flores, C. (2019). *Unlock 2* (2nd ed.). Cambridge University Press.
- Eiken. (n.d.). *Details of the Eiken CSE Score*. Retrieved December 11, 2024, from https://www.eiken.or.jp/eiken/en/eiken-tests/overview/eiken-cse_admission.html
- ETS. (n.d.). *Comparing TOEFL iBT Scores*. Retrieved December 11, 2024, from <https://www.ets.org/toefl/institutions/ibt/compare-scores.html#accordion-ff2afbe1ab-item-95a6512adf>
- MEXT. (2020). *Impact of COVID-19 pandemic on HE and MEXT's main countermeasures*.
https://www.mext.go.jp/en/content/20200707-mxt_kokusai_000005414_02.pdf
- O'Sullivan, B. (2020). *The comprehensive learning system: British Council perspectives on English language policy and education*. British Council.
https://www.britishcouncil.org/sites/default/files/the_comprehensive_learning_system_new_layout.pdf
- Tannenbaum, R. J., & Wylie, E. C. (2019). *Mapping the TOEIC tests on the common European framework of reference for languages*. ETS.
<https://www.ets.org/pdfs/toEIC/toEIC-mapping-cefr-reference.pdf>

Peer Feedback in Japanese EFL University Classes: Perspectives from Native English Instructors

Daniel Anderson, Asia University

Abstract

This article looks at peer feedback in Japanese university EFL classrooms from the viewpoint of the EFL instructor. Particularly, this article explores what EFL instructors perceive as benefits and setbacks of peer feedback and challenges EFL instructors encounter when implementing peer feedback in the classroom. Furthermore, this article seeks to illuminate ways to effectively implement peer feedback that may be useful for EFL instructors across Japan. To achieve these goals, I interviewed 10 university EFL instructors at a Japanese university in West Tokyo.

Introduction

In the field of language learning, classroom assessment has been a growing area of interest in the past few decades (Widiastuti et al., 2020). Sadullaeva et al. (2019) note that assessment is an integral component of the education cycle, not only because it confirms learning outcomes but also offers other functions including motivation for students to progress further. Two broad types of assessments found in EFL/ESL educational contexts are summative and formative assessments. Guado and Boersma (2018) state that summative assessment usually occurs at the end of an instructional course and documents the aggregate of learner achievements with the aim of summarizing overall learner accomplishments. Formative assessment, on the other hand, aims to promote feedback on learner performance to advance learning (Sadler, 1998). Furthermore, Guado and Boersma (2018) note that the main concern of formative assessment is neither to measure learner proficiency nor to verify learner achievements but rather to assist learner development through identification of points in need of improvement and the attendant actions needed for correction. Though both forms of assessment are important and work well together (Mashanian, 2019), there are numerous studies that point to the benefits of using formative assessment (Zeng & Huang, 2021). Glazer (2014) posits that summative assessments usually give students a score but offer no actual feedback, and therefore these types of assessments do not assist students in the learning process. Farhady and Selcuk (2022) note that formative assessments enable instructors to see what progress students have made, and this information enables them to adjust their instructional approach and learning materials as needed. One type of activity that falls under the umbrella of formative assessment and allows students to participate in the class is peer feedback (Saito, 2013). Wakabayashi (2008) notes that, with peer feedback, learners work collaboratively to exchange ideas and provide feedback on one another's writing. Saito (2013) describes peer feedback as a process where "students apply set criteria to the work of their peers in order to assess and provide feedback."

Literature Review

This section covers benefits of peer feedback for both students and teachers, and looks at some common drawbacks associated with peer feedback in Japanese and Asian EFL classroom settings.

Benefits of Peer Feedback in a Japanese EFL Setting

According to Saito (2013), studies have revealed various benefits of peer feedback for the ESL/EFL classroom. These benefits include an increase in student motivation (Hirose, 2014), an uptick on time spent for student-to-student interaction (De Guerrero & Villamil, 2000), an improvement in communicative ability (Mittan, 1989), self-awareness of one's own mistakes (De Grez et al., 2012), and a lightened workload (Topping, 2009).

Student Benefits

Through their study of two students giving mutual peer feedback, De Guerrero and Villamil (2000) noticed a scaffolding process naturally occur between the students that entailed an increase in student-to-student communication. Hirose (2014) also observed an uptick in student engagement time, but there is also evidence to suggest peer feedback can improve communicative ability. Mittan (1989) noted that through the peer feedback process, ESL students can tap into their communicative potential through language exchanges with other students. Through the process of peer feedback, some studies show that it is not only the receiver of the feedback that benefits. De Grez et al. (2012) posit that over the course of reviewing peer work, the reviewer is expected to establish a solid grasp of the expectations of the assignment, thereby scrutinizing their own work. Moreover, a study by Lundstrom and Baker (2009) compared the effects of giving versus receiving feedback and found that those giving feedback improved their writing more than those receiving the feedback.

Teacher Benefits

When learners take on the responsibility of giving feedback to other learners in the classroom, the teacher workload can decrease, which provides some relief because teachers are often inundated with work (Topping, 2009, as cited in Saito, 2013, p. 431). Additionally, Colpitts (2016) suggests that a notable benefit of implementing corrective peer feedback in the classroom is the time it saves for the teachers.

Drawbacks of Peer Feedback in a Japanese EFL Setting

Studies have shown that Chinese and Japanese students have a negative view of peer feedback (Mangelsdorf, 1992; Zhang, 1995). Two major issues with implementing peer feedback are allaying students' apprehension and guaranteeing consistency (Nilson, 2003). Another issue is that students are unfamiliar with the practice of giving and receiving peer feedback (Hirose, 2014).

One factor involved in student apprehension towards peer feedback is a lack of confidence. Saito (2013) conducted a study that looked at the perceptions of peer feedback of 46 first-year university students. The study involved discussions on various topics followed by peer feedback. Saito noted that some students believed they were unfit to provide feedback on their classmates' work. Baerischmidt (2013) similarly noted that students' belief that they themselves were unable to offer dependable feedback led to a student desire to do away with peer feedback.

Saito (2013) also stated that consistency of peer feedback could not always be guaranteed because some students felt obliged to give high assessments to those students with whom they had good relations. Additionally, insufficient training before implementing peer feedback led some to question the neutrality of the peer feedback they received (Saito, 2013).

Furthermore, the students' level of English or perceived level of English may adversely impact their experience with peer feedback. Yoshida (2008) observed that lower-level students can have difficulty understanding corrective feedback from higher level students. Lastly, peer feedback is not a common practice in Japan. Hirose (2014) notes that Japanese students are not used to sharing their work or providing feedback to other students. Colpitts (2016) conducted a study on 21 students enrolled in an English writing course at a Japanese university and found that the students preferred teacher feedback to peer feedback. Colpitts (2016) also notes that peer feedback can be challenging in Japan because the education system is set up around the notion of passive learning, meaning the teacher hands down knowledge to the students who receive it. This teacher-centered approach is the case not only for Japan but also for East Asia in general, where the teacher is viewed as an exemplar of conduct (Phong-Mai et al., 2020).

Research Questions

Though there has been a lot of research looking at student beliefs around peer feedback (Baierischmidt, 2013; Colpitts, 2016; De Grez et al., 2012; Mangelsdorf, 1992; Saito, 2013), this study seeks to illuminate some of the views of native English instructors who regularly implement peer feedback in the Japanese EFL classroom. The primary objective is to address the following research questions:

1. What are the perceived benefits of peer feedback in the EFL classroom according to EFL university instructors in Japan?

2. What are the perceived drawbacks of peer feedback in the EFL classroom according to EFL university instructors in Japan?
3. What are some recommended approaches, according to (the cohort of) EFL instructors in Japan, to effectively implement peer feedback in Japanese EFL classrooms?

Method

Participants and Institutional Background

In this study, a total of ten teachers participated. Among these teachers, nine were native English speakers, and one was a non-native speaker of English but had learned English from elementary school in the Philippines. The instructors all had at least a four-year degree, spanning fields in both the arts and sciences, and among them, three instructors held a master's degree in applied linguistics. Two of the instructors were female, while the rest of the instructors were male. The instructors ranged in age from early 30s to mid-50s. All instructors worked at the same university in Tokyo and taught first and second-year university students using the same textbooks for each grade. All their classes focused on communicative English without any writing components. It should be noted that peer feedback was a required element of the curriculum at the university, so all instructors implemented some degree of peer feedback in their lessons. Their class sizes ranged from eight to 25 students, with an average of 20 students per class.

Procedure

To understand instructors' beliefs about peer feedback in the EFL communicative-based classroom, I carried out an interview with 10 EFL teachers. Prior to the interviews, each instructor was requested to fill out a consent form that ensured they understood the reason for the interview, that the interview was being recorded, and that they would remain anonymous. The interviews proceeded only after the instructors signed the consent form. The interview consisted of 13 questions (see Appendix). This was a semi-structured interview as the questions were asked sequentially. However, there appeared to be a lot of overlap in the answers to the questions, and thus many participants answered questions in the early stages of the interview that were intended to be addressed later. Furthermore, there was a lot of tangential information brought up by the participants while they answered the original questions.

The interviews were all carried out on campus when the instructors had a break from lessons. The length of the interviews varied by the participant, with the shortest one lasting just 15 minutes and the longest one lasting 49 minutes. Once all the interview questions were answered, the participants were asked to give any final comments they may have had about using peer feedback in a Japanese university EFL setting.

Data Collection

The interviews were all recorded using an HP laptop. After all the interviews were finished, I ran the recordings through transcription software provided by the company Otter.ai. The transcriptions were mostly accurate, but I needed to go through each of the recordings manually to fix a few errors in each of the transcriptions. Notably, long pauses were tracked by the software as a separate answer to the question the respondents were originally answering, so these instances needed to be edited together to show continuity of the response.

Data Analysis

Once all the interviews were completely transcribed and edited, the answers were grouped together according to the following three categories: instructor-perceived benefits of peer feedback, instructor-perceived drawbacks of peer feedback, and instructor-recommended approaches (including any perceived pitfalls to avoid). These categories were then assigned the labels B (benefits), D (drawbacks), and RA (recommended approaches), respectively. Additionally, another category was created to include any other insights the instructors may have had. This category was labelled M for miscellaneous.

To compare the participating instructors' responses while keeping their names anonymous, each instructor was assigned a code of P plus a number from 1-10. For example, the first participating instructor was given the value P1. Accordingly, the second participating instructor was assigned the value of P2. This pattern continued until the last participating instructor, who was given the value P10. After all the instructors were given values, I was able to go through each interview and put the instructors' responses into one of the established categories and then compare the responses among the participants.

Results and Discussions

The first research question was as follows:

What are the perceived benefits of peer feedback in the EFL classroom according to EFL university instructors in Japan?

To answer this question, it should first be noted that all ten of the participating instructors remarked that there were at least some benefits of peer feedback in the Japanese university EFL classes. However, some participating instructors noted only minor benefits, while others said they noticed several benefits in the classroom. One benefit of peer feedback in the EFL classroom mentioned by some participating instructors was learner awareness of one's own mistakes when giving feedback. This metacognitive benefit was believed to help students make progress with their own language ability. For example, P1 (participating instructor 1) noted that peer feedback "makes students aware of what others are doing, which they are not doing, or what others are not doing, which they should be doing." P9 said that if a student notices others making some mistakes, they may become aware of those same mistakes for themselves. Furthermore, P9 gave the following example: "If a student notices another student making pauses in their speech, they might decide, 'I have to be careful not to make pauses in my speech'." P10 also mentioned that, when using peer feedback for speeches, the listener is observing the speaker, and this observation can help the listener notice any gaps in their own speech.

Another commonly reported benefit of peer feedback was that it can help build a harmonious classroom environment through pair and team interactions. P1 noted that the opportunities presented by peer feedback for students to talk with each other increases class rapport and helps students become more comfortable "putting themselves on the line" with other students. P4 noted the following: "I do think that one of the primary benefits of giving and receiving peer feedback is the ability to collaborate, not just in ideas, but in teamwork." P8 had the following to say about peer feedback in the EFL classroom: "I think it really increases positive atmosphere in the class, where you're really showing that, like, we're all in this together. Yes, and everybody has good things, everybody has things to improve."

A third benefit mentioned by some participating instructors is that peer feedback encourages active listening. P1 noted that, especially during speeches, it is good to see other students writing down feedback for the speaker because it shows they are actively listening, and this makes it less likely for students to tune out when their classmates are giving a presentation. P10 also noted that peer feedback was good for speeches because the students will actively listen to the speakers instead of just waiting for their turn to speak.

Other independently (not shared with other participating instructors) observed benefits of peer feedback include time management for classes that are too big for the teacher to give everyone feedback, practice developing the skill of giving advice, and experience in dealing with criticism.

These results show that some of the observations made by the participating instructors tie into the literature on the benefits of peer feedback in ESL/EFL contexts. For example, the observation of metacognitive benefits aligns with the 2009 study by Lundstrom and Baker that showed greater benefits for the student giving feedback than for the one receiving it. Furthermore, the observation that peer feedback can help students create a rapport likely stems in part from the time the students are communicating with each other. Hirose (2014) noted the increase in student-to-student engagements with each other as a benefit of peer feedback. Finally, although only noted by one participating instructor, the observation that peer feedback can help the workload of a teacher with a large class coincides with the study by Topping (2009) that showed peer feedback can help alleviate work for the busy teacher.

The second research question was as follows:

What are the perceived drawbacks of peer feedback in the EFL classroom according to EFL university instructors in Japan?

As with the benefits of peer feedback, all participating instructors agreed there were drawbacks as well. Again, some teachers perceived only minimal drawbacks, while others felt there were several drawbacks. One common drawback noted by participating instructors was that most students were unfamiliar with peer feedback prior to taking the course, which made it a challenge for students to get comfortable using it. Some instructors also thought that cultural factors may present challenges in peer feedback among Japanese university students. P1 noted that, for peer feedback, “students were more comfortable with a teacher-student relationship.” P10 also noted that it is a challenge in “getting [students] over that kind of student-teacher one-way relationship that they’re used to where they are just passive participants in the classroom.” P9 said, “The Japanese phrase 出る杭は打たれる (The nail that stands out gets hammered down) comes to mind. And many Japanese students don’t like to stand out or be pointed out.”

Another challenge mentioned for implementing peer feedback in the Japanese EFL classroom is having students give critical feedback. P2 noted, “They are hesitant to criticize each other in any way, especially the hard ways.” P5 said, “They are quite nice to each other generally, and they’re often not very critical. So perhaps they are not really giving their

partners anything they can necessarily work on.” Furthermore, P9 noted, “If a student tells their partner ‘you made a mistake,’ it just causes poor feelings.”

A further challenge of peer feedback noted by some participating instructors is that students may give feedback that is too general. With regards to student speeches in class, P5 noted that feedback such as “loud voice” or “big smile” is too general and clichéd to be of any use to the receiver. P3 also gave student speeches as an example and said sometimes students would just say “More eye contact” for feedback. P3 believed that this is probably due to a lack of effort, noting “There are sometimes, some students that give you, they’re just going through the motions. So, it’s not really fair to the students who are trying their best.”

Some of these observations from participating instructors match up with the existing literature. For example, Hirose (2014) noted that Japanese students were unfamiliar with giving and receiving peer feedback. Furthermore, the observation that students are more comfortable with the teacher-led classroom, where feedback is provided by the teacher, corresponds with different studies on peer feedback in Asia (Colpitts, 2016; Mangelsdorf, 1992; Phong-Mai et al., 2020).

The third research question was as follows:

What are some recommended approaches, according to (the cohort of) EFL instructors in Japan, to effectively implement peer feedback in Japanese EFL classrooms?

The participating instructors seemed to agree on some basic approaches to carrying out peer feedback in the classroom. One simple suggestion mentioned by different participating instructors is to make the criteria of what is being evaluated clear. P1 stated that, for speeches, students should focus their feedback on “verbal delivery, nonverbal delivery, speech message, and speech structure.” For discussions, the focus should be put on “active communication, reasoning, and development of the topic.” P3 mentioned the importance of having students know what they are being evaluated on and in turn what they should be thinking about for the peer feedback, noting “I try to get them to focus on feedback for graded points, which will hopefully lead them to realizing what they need to do to improve their score.” Furthermore, P10 said, “If there’s evaluation criteria, I won’t give them all the criteria, but I’ll give them the categories and will do an activity, like around boarding, to elicit from them, to have them write on the board what they think the criteria are and then go over that with them.”

Another suggestion by some participating instructors was to have the students be specific in their feedback. P1 noted that simply expressing “you need to improve your

speaking” to another student is not very useful. Instead, expressing “you could improve your ‘th’ pronunciation” would be more useful for the receiver. P7 noted that giving examples was important for the receiver, saying “‘I like your message’ is not clear. What was it about the message you liked? Give an example of what you liked about the message.” P9 said that using the word “because” was helpful in getting students to be more specific, such as in “Something was good because blank” or “Something needs work because blank.”

A further suggestion for peer feedback was to include both positive and negative (points to improve) feedback. P2 noted that a mix of comments was important, stating “We want to try and make clear that both positive and improvement [comments] are necessary. And they shouldn’t focus on just the positives.” P4 stated the following: “My number one rule is to keep peer feedback balanced. One strength, one improvement point. Or two strengths and two improvement points.” P10 noted that although he starts off with a strong point and then a point to improve when modeling how to give peer feedback, “I don’t have a rigid rule that says they need to say a good point first, only that they need a good point and a point to improve.” Likewise, P8 noted that students should have “one positive and one improvement point. I don’t think the order is important.” P8 further noted that after the students give a positive point and an improvement point, he asks the students giving feedback to ask one question to the student receiving feedback (based on the assessed content) to help further communication.

Question 12 of the survey asked if the participating instructors think templates are useful for setting up peer feedback and, if so, what words or phrases did they use. Most of the participating instructors said that some kind of template was useful to model peer feedback, especially in the beginning of the course when introducing peer feedback to the students. Many instructors encouraged their students to use phrases that they [instructors] were provided by the head instructor. For giving a strong point, the phrase “I liked that you _____” was used by most instructors. For example, in a speech, a student could say, “I liked that you made eye contact with the audience.” For giving a point to improve, the phrase “Next time you could _____” was also used by most of the instructors, such as “Next time you could speak in a louder voice.”

Other recommendations offered by participating instructors included starting peer feedback in the course as soon as possible, staying consistent with the format of peer feedback, having students write down the feedback before they tell their partner, and having students work in groups to decide peer feedback through a collaborative process.

Some of the suggestions by the participating instructors tie into the literature of best practices for peer feedback. Black and William (1998, as cited in Saito, 2013) stressed that giving specific information is a vital component of peer feedback. They noted that giving obscure feedback, such as “You did a good job!,” was not useful to the receiver. Mangelsdorf (1992) posited that modeling the feedback with the students is an effective way to get the students comfortable using peer feedback. This aligned with the belief of participating instructors that using set phrases, or templates, was a good way to model the feedback. Furthermore, Mangelsdorf (1992) stated that it is important to make sure the students know what criteria they should be looking for when providing peer feedback.

Pedagogical Implications

Looking through the results, there are some practical approaches in the Japanese EFL classroom that teachers can employ to help ensure smooth implementation of peer feedback. A few of these approaches highlighted in the study reinforce some of the literature on best practices for peer feedback, while other approaches present ideas less commonly known.

Some recommendations from the instructors in this study that are backed by existing literature include giving balanced feedback that is both positive and critical, modeling the feedback to show students how to carry out peer feedback, and avoiding general feedback while encouraging feedback that is specific. For example, when giving feedback for a speech, avoid saying “Great job” and use something more specific, such as “Your example in the introduction was very clear.”

Looking at some more novel approaches presented in this study, teachers could benefit from introducing practices as follows. When students give peer feedback, especially for speeches, have them write down the feedback first. This can help ensure that they are actively listening and have something to show the person they gave feedback to in addition to the verbal feedback. Another idea is to ask a question when following up with feedback. This is useful for discussions and speeches because it ensures that the person giving feedback is paying attention to the person speaking, and it allows for more interaction time between the pair or group. Though templates were mentioned as a good idea that ties into the peer feedback literature, the following two phrases were recommended by several teachers to use for giving positive and critical feedback. For positive feedback, teachers can introduce the phrase “I like that you ____.” For an area that needs improvement, teachers can introduce the phrase “Next time you could ____.” Finally, carrying out peer feedback as early in the

semester as possible is a good way to give students early exposure to the process, which can give them more confidence as the semester progresses.

Study Limitations

This study was limited to a cohort of only 10 instructors. Furthermore, all 10 instructors were from the same university, which meant they adhered to the same rules and expectations of that university. As all instructors were required to carry out peer feedback, they may have beliefs about peer feedback not shared with teachers who have never used peer feedback or use it sparingly. A future study that involves a larger cohort of teachers across different universities could provide better insight into what the larger EFL instructor community in Japan feels about peer feedback and what additional insights they may have in implementing peer feedback in class. Another limiting factor of the study was the broad nature of the questions for such a small cohort from the same university. Future studies of similar scale could be more specific, asking instructors what they do at their particular institution (without naming the institution). However, if a large-scale study is undertaken at different universities, the broad nature of the questions could be suitable.

Conclusion

In this study, 10 university EFL instructors were interviewed on their views of peer feedback with the goal of gaining insight into the benefits, disadvantages, and best practices to carry out peer feedback in the Japanese university EFL classroom. In summary, participating instructors noted that the benefits of peer feedback include metacognition of one's own language on the part of the learner giving feedback, encouragement of a cohesive and harmonious classroom, and promotion of active listening. In contrast, participating instructors said that challenges of peer feedback included unfamiliarity of peer feedback among Japanese university students, reluctance to give critical feedback, and the tendency of some students to give feedback that is too general and thus not beneficial to the receiver. Finally, participating instructors noted that best practices of implementing peer feedback included making sure students are aware of the criteria used for the feedback, having students give specific feedback, mixing feedback with positive points and points to improve, and modeling the feedback with set words and phrases. Taking into consideration the insights and best practices offered by the participating instructors can serve as a valuable tool for EFL teachers in Japan to include in their repertoire of classroom approaches.

References

- Baierschmidt, J. (2013). Japanese ESL learner attitudes towards peer feedback. *The Journal of Kanda University of International Studies*, 24, 101–114.
- Black, P., & William, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80(2), 139–148.
- Colpitts, B. D. F. (2016). Japanese students' perceptions of peer corrective feedback in an EFL classroom. *Humanities Series*, 49, 345–358.
- De Grez, L., Valcke, M., & Roozen, I. (2012). How effective are self- and peer assessment of oral presentation skills compared with teachers' assessment? *Active Learning in Higher Education*, 13, 129–142. <http://doi.org/10.1177/1469787412441284>
- De Guerrero, M. C. M., & Villamil, O. S. (2000). Activating the ZPD: Mutual scaffolding in L2 peer revision. *The Modern Language Journal*, 84, 51–68.
- Farhardy, H., & Selcuck, M. (2022). Classroom-based diagnostic assessment practices of EFL instructors. *Iranian Journal of Language Teaching Research*, 10(2), 77–94.
- Guadu, Z. B., & Boersma, E. J. (2018). EFL instructors' beliefs and practices of formative assessment in teaching writing. *Journal of Language Teaching and Research*, 9(1), 42–50. <http://dx.doi.org/10.17507/jltr.0901.06>
- Hirose, K. (2014). Written and spoken peer feedback in EFL writing instruction: A case with Japanese undergraduates. *The Journal of the Faculty of Foreign Studies, Aichi Prefectural University*, 46, 83–110.
- Lundstrom, K., & Baker, W. (2009). To give is better than to receive: The benefits of peer review to the reviewer's own writing. *Journal of Second Language Writing*, 18, 30–43.
- Mangelsdorf, K. (1992). Peer reviews in the ESL composition classroom: What do students think? *ELT Journal*, 46(3), 274–284.
- Mittan, R. (1989). The peer review process: Harnessing students' communicative power. In D. Johnson & D. Roen (Eds.), *Richness in writing: Empowering ESL students*, 207–219.
- Nilson, L. B. (2003). Improving student peer feedback. *College Teaching*, 5, 34–38.
- Park, J. (2018). Effectiveness of teacher and peer feedback: Through the lens of Korean Tertiary Writing Classroom. *The Journal of Asia TEFL*, 15(2), 429–444.
- Phong-Mai, N., Terlouw, C., & Pilot, A. (2005). Cooperative learning vs. Confucian heritage culture's collectivism: Confrontation to reveal some cultural conflicts and mismatch. *Asia Europe Journal*, 3(3), 403–419.

- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18, 119–144.
- Sadullaeva, S., Arustamyan, Y., & Sadullaeva, N. (2019). Shifting the assessment paradigm from knowledge to skills: Implementation of new appraisal procedures in EFL classes in Uzbekistan. *International Journal of Innovative Technology and Exploring Engineering*, 9(1), 4943–4952.
- Saito, Y. (2013). The value of peer feedback in English discussion classes. In N. Sonda & A. Krause (Eds.), *JALT 2012 Conference Proceedings*. Tokyo: JALT.
- Topping, K. (2009). Peer feedback. *Theory Into Practice*, 48, 20–27.
- Widiastuti, I. A. M. S., Mukminatien, N., Prayogo, J. A., & Irawati, E. (2020). Dissonances between teachers' beliefs and practices of formative assessment in EFL classes. *International Journal of Instruction*, 13(1), 71–84.
- Yoshida, R. (2008). Learners' perception of corrective feedback in pair work. *Foreign Language Annals*, 41(3), 525–541.
- Zhang, S. (1995). Reexamining the affective advantage of peer feedback in the ESL writing class. *Journal of Second Language Writing*, 4, 209–22.

Appendix

A Survey on Peer feedback

- 1) How often do you use peer feedback in class?
- 2) Did you have any experience having students give peer feedback before the current course?
- 3) Do you think your students have had experience with peer feedback (either giving or receiving) before taking your class? Roughly what percentage of the university students you teach do you believe have experience with using peer feedback?
- 4) When do you have students provide peer feedback?
- 5) How much time do you allot to the students for peer feedback?
- 6) What areas of assessment do students provide feedback on?
- 7) What do you think are some of the benefits for both the students giving and receiving peer feedback?
- 8) What do you think are some of the drawbacks for both the students giving and receiving peer feedback?
- 9) What are some challenges of implementing peer feedback in a Japanese University EFL setting?
- 10) Do you think the students value peer and teacher feedback equally? Why or why not?
- 11) Do you have any guidelines for the students when they provide peer feedback? If so, what are they?
- 12) Do you think it is helpful to provide the students with a template of expressions or words to use for giving feedback? If so, are there any phrases that you often use? If not, why not?
- 13) What advice would you give to EFL teachers implementing peer feedback in a Japanese University classroom setting?

Using Gamification to Incentivise Students in Mandatory TOEIC Classes

Denver Beirne, Asia University

Abstract

Gamification techniques have become increasingly popular in language learning classrooms. This interest is with good reason, as there is substantial evidence that the practice can increase student engagement and participation. Inspired by these findings, this study introduced a gamified environment into mandatory TOEIC lessons to leverage this ability to increase student motivation. Several games were integrated with the textbook activities to supplement conventional instructional methods. Students were then surveyed on their attitudes to the game activities. The paper presents the results of this survey of 76 participants. The study found that over 90% of students preferred this learning style to traditional learning. However, a significant minority of students reported being motivated by the rewards associated with games rather than improvements in language skills.

Introduction

This study was conducted at a private university where students must complete mandatory English classes in their first year. By the time the participants reached university, they were still at beginner or elementary level of English, despite having already completed several years of compulsory English lessons. During an interview activity at the beginning of the year, most of the students stated that they did not like English. Hence, the participants had low motivation. Given this situation, the study introduced gamification techniques to increase students' motivation. The project's success would be judged by whether or not the students felt incentivised by the activities. This paper will describe the study methods and activities used, and it will also present the result of a survey on student attitudes towards the game activities. Firstly, to contextualise the findings, the following section will review the existing literature on gamification in the language learning classroom.

Gamification in the Language Learning Classroom

Gamification can be defined as the use of game mechanics in non-game contexts (Deterding et al., 2011; Werbach & Hunter, 2012). In the language learning context, this often translates to using tools such as points, leaderboards and badges to reward competitive challenges. As Flores (2015) states, "The main objective of Gamification is to increase participation and motivate users through the use of game elements..." (p. 37). The goal is not to recreate a (video) game in the classroom but to use some game mechanics in language activities to "...encourage and reward behaviors that support learning and foster productive social interactions" (Hung, 2017, p. 58).

Studies on gamification have found that the giving of points was the most popular gamification tool (Alomari et al., 2019; Antonaci et al., 2019), with badges/awards and leaderboards as the next most popular elements. There is a preponderance of evidence suggesting that this popularity is for good reason. Firstly, the competition and fun fostered by a points system can increase student motivation, participation and emotional states (Antonaci, 2019; Mee et al., 2020; Yaccob et al., 2022). Game elements such as points, supported by a leaderboard, can also incentivise students by allowing them to compare their relative performance against their classmates' (de Byl, 2013). Furthermore, team activities can promote cooperation as well as competition, as the groups need to collaborate to accumulate points or rewards (de Byl, 2013; Hung, 2017; Matsumoto, 1989).

Conversely, there is a danger that some students can become demotivated by failure in competitive endeavours, so this prospect must also be considered. However, it has been found that (occasional) classroom failure can also be a motivational factor for some students (Dickinson, 1995; Matsumoto, 1989). Nonetheless, a degree of caution is needed, as adverse effects of gamification have been reported. For example, some students have described how gamification can become too complicated or competitive (Hung, 2017).

The most extensive meta-analyses of the literature (Dehghanzadeh et al., 2019; Laura-De La Cruz et al., 2022; Nurlery, 2024; Rahmani, 2020) concluded that while there is evidence for increased motivation, there is a lack of robust evidence to support learning outcomes that exceed traditional methods. For example, Buckley and Doyle (2014), and Harmilawati et al. (2023) have claimed that game-based learning increased language skills. These claims are valid. However, neither research group compared these positive results against a control group that learned through non-game-based methods. Therefore, it is not possible to assert that the increase observed outperforms traditional methods. Meanwhile, Tamayo et al. (2023) also discovered an improvement in language skills for a group of students using gamified activities. This study did create a control group who learned using non-game-based methods and found no statistically significant difference in the improvement between the game-based group and non-game-based-group.

Therefore, in sum, the literature indicates that games can increase motivation and may improve learners' performance or participation, especially in specific tasks. In addition, there is some indication that gamification is an effective method to improve students' language skills, yet not necessarily better than traditional methods. Thus, teachers should not expect the observed student enjoyment necessarily to translate to lasting improvements in language skills. With this literature in mind, this study set out to incentivise students on a mandated TOEIC course and report on the findings.

Research Questions

The study aimed to investigate whether gamified tasks could improve students' motivation and participation in TOEIC lessons. The goal was not to attempt to measure these variables objectively but to gain an insight into students' perceptions of the learning methods. As a result, the following research questions were created:

1. Do students enjoy the game activities?

2. Do students believe the game activities are motivational?
3. Do students believe the game activities improve their language skills?

Methodology

Participants

There were 76 participants at a beginner or elementary level of English (CFER A1-A2), majoring in Data Studies, Law or Hospitality. All the participants were enrolled in one of five mandatory Freshman English classes. Each class contained 16-20 members.

Procedure

The mandatory Freshman English course requires that half of the lessons are allocated to TOEIC studies. (The Freshman English course overall followed a regular pattern of alternating between one week of TOEIC studies and one week following a different textbook.) For these particular students, the TOEIC textbook was ABAX's *TOEIC Skills 1* (Graham-Marr et al., 2016). During the semester, the TOEIC lessons followed this textbook closely, but most of the activities were gamified in some manner. Students were surveyed at the end of the semester about their attitudes to the game activities, with a focus on the TOEIC materials.

The survey data was collected using a Google Form questionnaire (reproduced in the Appendix). All questions were provided in English and Japanese. The survey used a mix of quantitative and qualitative methods to provide a deeper understanding of students' attitudes. There were eight closed questions (categorical multiple-choice questions and a Likert scale) and one open-ended question where students could comment at greater length. Students were permitted to write in English or Japanese.

Games, Activities and Gamification Techniques

Points

The course used a class points system. This process is explained in detail in Beirne (2023); however, a brief overview will be provided for context. Students were divided into groups for each lesson. Points earned for each activity were written on the board. At the end of the lesson, students added their class points to a Google Sheet. At the end of the semester, students' points were totalled and converted to percentages, which formed part of their class participation grades. This process aimed to provide a competitive focus to the lessons and

course by enabling students to see their scores build in relation to their efforts and compare themselves against their classmates. The reward scheme for most of the activities is shown below:

- 1st place – 10 points
- 2nd place – 7 points
- 3rd place – 5 points
- 4th place – 3 points

The following section will now focus on some of the specific activities used on the course. It is important to stress that these activities were combined with more traditional methods in which grammar, vocabulary and conceptual information was explained or illustrated more explicitly before or after activities.

Vocabulary Activities

Quizlet Live

Students use a QR code to join a vocabulary-learning game within the Quizlet flashcard application and are sorted into teams. The teams match the vocabulary item to its meaning from four multiple-choice options. The winner is the first team to answer an unbroken chain of 12 questions.

Blackboard Karuta

The teacher prints a selection of images of the week's vocabulary and places them on the board with magnets. Teams line up in front of the board. The first member of each team is given a flyswatter. The teacher says one of the vocabulary items, and the first student to hit that item receives points for their team. The difficulty can be increased by saying the definition rather than the vocabulary item.

Hot seat / Taboo / Circumlocution

The teacher prepares several images of the week's vocabulary items on presentation slides. One member of each group sits on a chair (the hot seat) with their back to the projector or TV. The other team members gather around their hot seat. The teacher shows the first image on the screen. Students describe the vocabulary item but cannot say any part of the

vocabulary word. The first hot seat member to guess the word wins points for their team. Lower-level students can also use gestures.

Vocabulary in Context

Competitive Gap-Fill

Students use one textbook or worksheet per group. The first student in a group adds the first missing word to the gap-fill, then passes the textbook/worksheet to the next student in the group, who adds the next word. The first team to finish all the sentences correctly is the winner.

Sentence Writing Gap-Fill

This version of the game gives students extra writing practice. Instead of using a textbook or worksheet, teams write on a blank piece of paper. The first student in the group writes the first full sentence of the gap-fill, including the missing word. The next student writes the next sentence. This is repeated until all the sentences have been completed. One variation of the game has students write the sentences on the board. The first member of each team writes the first sentence, including the missing word, as fast as they can. The first student to finish the sentence receives points for their team. The teams rotate, and the next member of each team writes the next sentence.

Readings

Running Dictation

The text is placed at the back of the classroom. One student is designated as the writer for each team. The other students take turns checking and memorising as much of the text as possible, then dictate this to the writer. Only one team member can check the text at a time. The first team to complete the text correctly receives maximum points.

Telephone Game

One member of each team is designated as the writer and sits at the front desk. Students make a line behind their writer, and the text is placed at the back of the class. Only the team member closest to the text can check the reading. The student closest to the text whispers part of the reading to the next team member and this continues until the message

reaches the writer. This process is repeated until the text is completed, and points are allocated based on the teams' finishing positions.

Blackboard Telephone Game

The process is the same as the regular version, except the writer is positioned in front of the board, where the text will be written. This game works best by having teams compete to finish just one sentence at a time, then resetting and cleaning the board after each round.

Speed Dictation

Each team uses one textbook to read and one sheet of paper to write. The teacher designates one student as the writer. The other students take turns to dictate one sentence each until the text is completed.

Sentence Jumble

The teacher prints out a text, cuts it up into its constituent sentences, mixes up the strips of paper and gives one set of strips to each team. The same effect can be achieved digitally by using Google Slides, Sheets, or Documents, as multiple users can edit a file simultaneously: In a single document, the teacher rearranges the sentences in the reading. He or she makes one copy for each team and labels them appropriately. Each team competes to reassemble the reading in the correct order.

Answering the Reading Questions

After completing the activities, the readings can be reviewed with the class as necessary. Students answer the questions individually or as a pair/group activity. (Usually, students answer some of the questions as a group activity and some individually.) Points can be awarded for correct answers in either case. The individual practice mimics the test situation, but the group practice gives lower-level students some shared experience from which to draw.

Listening Activities

Students in each team are allocated team member numbers using *janken* (rock-paper-scissors). The teacher plays the first three listening problems. Team Member 1 writes the answers on paper. After an allotted deliberation time, all the writers transfer their answers to

the board at the same time. The teacher reviews the answers—explaining vocabulary, phrasing, usage or grammar as necessary—and then awards points for correct answers. Team Member 2 becomes the writer for the next set of listening questions, and so on. As an alternative to this method, all the members of each team can congregate at the board and write the answers directly on the board. As with the reading questions, students usually complete some questions individually and some as a group activity.

Additional points can be awarded for the above game in various ways. One concrete example is a missing-picture challenge. Teams are shown a slide containing pictures of some of the week's vocabulary. Then, the students are shown another version of the slide with one picture removed. The first team to say the missing item in English receives double points for the three listening questions that follow.

Miscellaneous Activities

Categorisation Activities

Some sections of the textbook ask students to add words to a category such as word type (e.g., verbs, nouns and adjectives). These activities can be completed competitively in pairs or groups, with points given based on how fast each activity is completed. As an extension, students can be asked to create original, grammatically correct sentences for each word. This activity can also be conducted competitively, using the same pattern as the sentence writing gap-fill.

Expanding Sentence / One-Word Sentence

This activity can be used to introduce the theme of a chapter. The teacher gives each team a sentence starter. For example, if the theme is healthcare, the sentence starter could be, "Last time I went to the hospital, ..." Students line up at the board and take turns to add just one word to the sentence. Each team strives to create a single grammatical sentence within a time limit. There is an emphasis on the correct usage of conjunctions and sentence structure/sentence order, and teams are deducted one point for each spelling or grammatical error.

Grammar Correction

Section B5 of *TOEIC Skills 1 (Focus on Grammar)* depicts five correct and five incorrect sentences. Students are required to find and correct the incorrect sentences. The

teacher sets a time limit for student teams to complete the first half of this exercise (Questions 1-5). The teacher reviews the answers as a class and gives each team a point for each sentence they identify correctly. Then the teacher gives each team the opportunity to gain bonus points if they can correct the erroneous sentences. Teams then complete the second half of the exercise (Questions 6-10) and receive double points for each correct answer. Alternatively, teams could gamble the points they earned in Questions 1-5, which adds the option of turning this activity into a *Grammar Auction*.

Grammar Correction Race

The teacher prepares strips of paper that each contains a different sentence from Section B5. The captain of each team takes one strip of paper, deliberates with their team, and decides whether the sentence is correct or incorrect. If the team believe the sentence is incorrect, they must also provide the correction. If their answer is confirmed by the teacher, they can take a new sentence. The teams race to finish all ten sentences.

Blackboard Spelling Bee

Teams line up at the board. The first student in each line is designated the writer. The teacher says a word from the week's vocabulary. The first team to write the word on the board with the correct spelling receives points. Students rotate, and the next student in line becomes the writer. The spellings progress in difficulty with each round, so the points awarded also increase with each word. This activity can be effective as a review of the week's vocabulary and an enjoyable task to finish TOEIC week.

Results and Analysis

The paper will now analyse the survey results to examine the students' attitudes towards using these games and activities.

Quantitative Results

Figure 1. Question 1: *Please rate how strongly you enjoy playing games in class activities.* クラス活動の中でのゲームにどのくらい楽しめたかを評価してください。

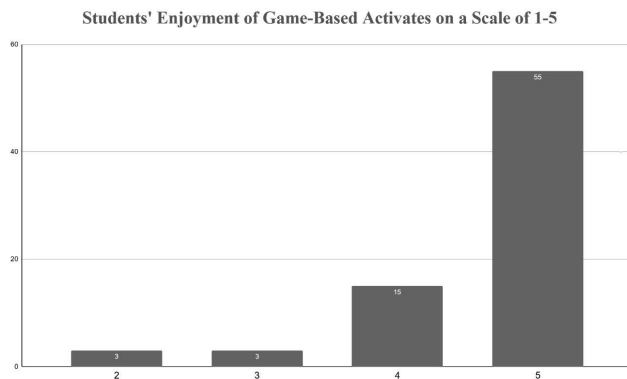


Figure 2. Question 2: *Do you feel you learn more in a game-based activity or more in an activity that is non-game-based (worksheet or interview, for example)?* ゲーム重視のクラス活動でより学びを得ることが出来ますか、それともゲーム以外の活動の方がそうだと感じますか？(例えば、ワークシートやインタビューなどの活動)

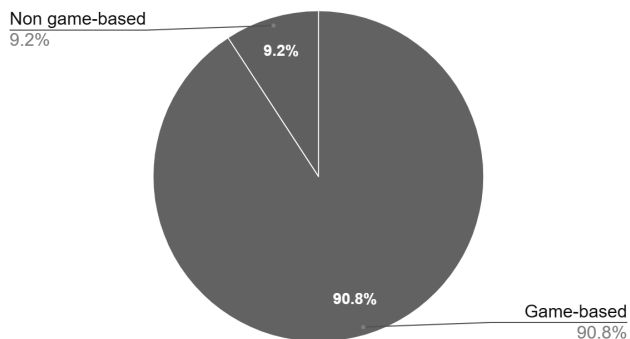


Figure 3. Question 3: *Do you feel more motivated by a game-based activity or more motivated by an activity that is non-game-based (worksheet, or interview, for example)?*

ゲーム重視の活動に対してモチベーションを感じますか、それともゲーム以外の活動の方がモチベーションになりますか？

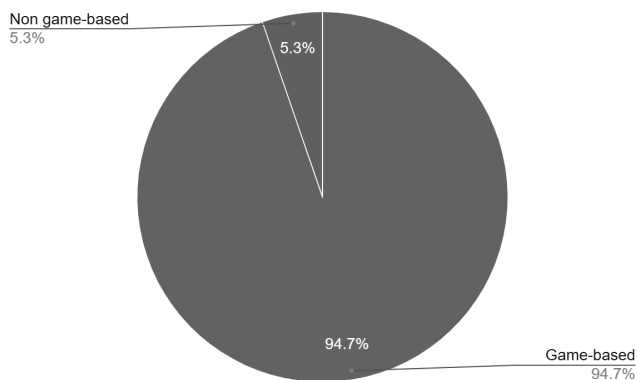
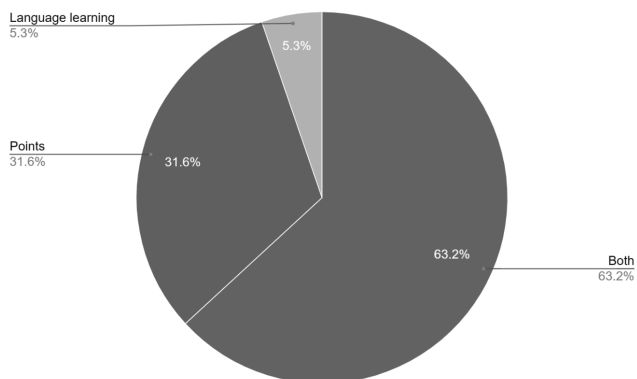


Figure 4. Question 4: *When you play a game in class, are you focused on getting the points or language learning?* クラスでゲームをしている際、ポイントを取ることに集中していますか、それとも言語を学ぶことに集中していますか？



Question 1 demonstrates that the vast majority of learners enjoyed the activities. The responses show that (55 out of 76) chose the highest rating of 5 (I really like it). Moreover, 70 of the 76 students (92%) chose one of the two highest ratings of 4 (I like it a little) or 5 (I really like it), while three participants logged a score of 3 (sometimes I like it, sometimes I don't like it). Only three students scored the activities negatively, choosing 2 (I dislike it a little), and no students chose 1 (I really dislike it). The results gave an average enjoyment score of 4.61 out of 5.

Questions 2 and 3 showed a clear preference for game-based activities, with around 91% believing they learned more through games and 94.7% stating that they felt motivated by the games. Question 4 highlights the issue of what actually motivates the students. Just 5% responded that language learning was their primary objective in these activities, while 32% directly stated that the awards were the key motivator. The majority (63%), though, claim to be motivated by both. However, it is unclear to what extent students in this group actually favoured each of these options and what effect this ultimately had on learning outcomes.

Taken together, these responses provide positive answers to all the research questions, as shown below:

1. Do students enjoy the game activities?
 - a. Yes. 91% of students answered "yes."
2. Do students believe the game activities are motivational?
 - a. Yes. 97.4% of students answered "yes."
3. Do students believe the game activities improve their language skills?
 - a. Yes. 91% of students answered "yes."

The students expressed a strong preference for learning through games, and they perceived them to be motivating. However, only 5% of students reported that language learning was their primary focus in these activities. Conversely, when this is added to the percentage of students who claimed to prioritise both points and learning, a healthy majority of 68% is reached. However, this still leaves over a third of students who are only focused on earning points. This finding reflects the debate in the literature about what underpins the incentivisation observed in game-based learning. It has been argued that extrinsic rewards, such as those given in gamification activities, could reduce intrinsic motivation (students' internally driven motivation) if not handled carefully (Deci et al., 1999; Hanus & Fox, 2015; Pulfrey et al., 2013). The reasoning asserts that if a student is given awards for a task they would otherwise enjoy, the focus might shift from the learning process to the reward.

The results of Question 4 indicate that a sizable minority of students are motivated by rewards. However, responses to the other questions suggest that these students would not otherwise enjoy these TOEIC activities. Therefore, it is not clear that their motivation has shifted from intrinsic to extrinsic, as, given their stated dislike of English, it is unlikely that they had intrinsic motivation in the first place. Furthermore, other scholars counter the idea that intrinsic motivation is damaged by gamification (Mekler et al., 2017; Rahmani, 2020). Next, the paper will turn to the TOEIC-specific questions.

Figure 5. Question 5: *Has the use of games and teamwork made TOEIC classes more enjoyable?* ゲームやチームワークを通して、TOEICのクラスはもっと楽しいものになりますか？

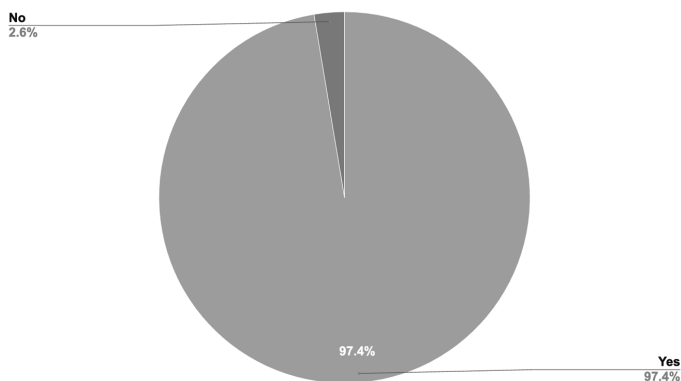


Figure 6. Question 6: *Do you believe you have learned useful language and communication skills through the games and teamwork used in TOEIC classes?* TOEIC のクラスで、ゲームやチームワークを通して有用な言語やコミュニケーション能力を学べたと思いますか？

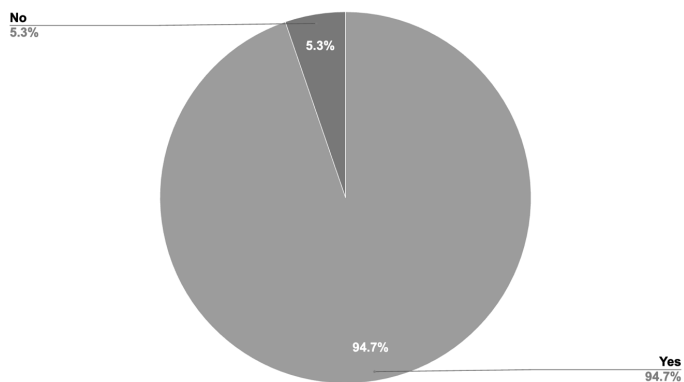


Figure 7. Question 7: *Do you believe your TOEIC skills have improved as a result of the games and teamwork used in classes?* クラス内でのゲームやチームワークを通してあなたの TOEIC スキルが良くなったと思いますか？

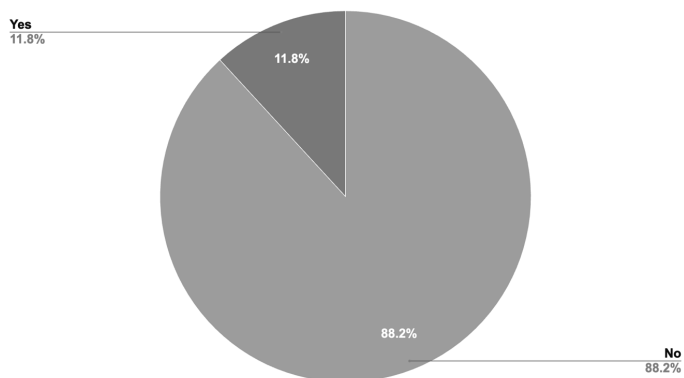
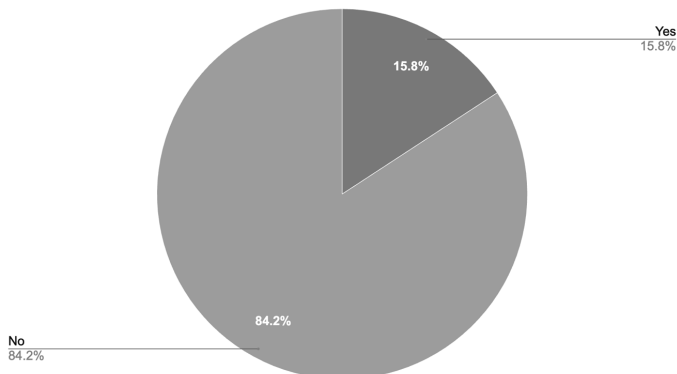


Figure 8. Question 8: *Has your TOEIC score improved over the past year?* これまでと比べ、あなたの TOEIC の結果は良くなりましたか？



These TOEIC-specific questions echo the previous results, with students overwhelmingly enjoying the activities (97.4%) and believing the tasks have improved their general communication skills (94.7%) and TOEIC language skills (88.2%). It must be emphasised that these are student perceptions rather than objective assessments of improvement. Nevertheless, this could have important implications for students' motivation.

The responses to the final question raise a number of important points. Firstly, students' TOEIC results were not monitored, as it was not the focus of the study, so these claims could not be verified. Secondly, even if 84% of students had increased their TOEIC scores, it would only show a correlation between the gamified activities, rather than a causal link.

However, Question 8 provides interesting insight into student perception. When Questions 7 and 8 are compared, a clear difference exists between student perception and student results (if self-report could be verified). While 88% of students believe the game activities improved their TOEIC skills, only 84% claim their TOEIC score had increased. This is a slight difference, yet it illustrates how perceived outcomes can diverge from assessed abilities. This difference may hint at why perceived increases in engagement have not been shown to produce results that outperform traditional methods.

Qualitative Results

Question 9: What is your opinion about using games in the classroom? Please write a few sentences about your views. クラス内でゲーム活動を用いることに対してどう思

いますか？あなたの意見や考えを文章でご回答ください。

As with the closed questions, the answers students posted in this section were overwhelmingly positive, with many focusing on how the enjoyment of the games increased their motivation. Out of the 76 comments, 33 used the word “good,” while 16 featured the words “fun” and “enjoy.” In addition, 18 students reported that the activities increased their motivation. The following is a typical comment: “It motivated me because game activities made the class fun.” This statement says nothing about long-term motivation; however, student engagement was clearly observable in the classroom, as illustrated by the comment, “It was pretty hard because almost all English classes were during first period, but I enjoyed learning with game activities.”

Several students made comments that indirectly reference motivation, with seven students describing the games as “active” and six students expressing their ability to participate positively. For example, one student commented, “It allows for enjoyable learning, so I wanted to actively participate in your class.” Students also mentioned the relationship-building that the teamwork encouraged: “A good way to make friend relationships [*sic*]. I enjoyed” and “Better than usual class without getting bored [*sic*]. It allows us to interact with classmates and become friends.”

In total, there were four negative comments recorded in this section. About game-based learning, one student stated, “It is good once in a while.” One might conclude that games are used too much for this student’s liking, although it is difficult to know precisely what is meant. Another student answered, “I was able to feel familiar with English. However, I sometimes concentrated on getting more points than learning English.” This statement echoes the discussion around Question 4 and the concern in the literature about intrinsic versus extrinsic motivation.

Two students offered more thoughtful and in-depth comments, with one focusing on how the allocation of team members can affect team points:

Game activities is increases [*sic*] student motivation. However, there were times when I prioritised getting points rather than improving my English skills. As in examples of English sentences that use “and” a lot. Also, in Quizlet, students were divided into those who studied and those who did not study at all. It was as if the

winners and losers were determined by team division. But I really enjoyed this past year. thank *[sic]* you very much.

The quote offers an interesting view of how this student feels about the fairness of team activities. The division of team members is an issue that can be challenging for teachers, and perhaps there is no ideal solution. In these lessons, teams were determined randomly, with the view that, over the course of the semester, randomness would even out any imbalances that occurred in a single lesson or activity. Even if teams are carefully balanced, as best a teacher can, there will always be students who study harder than others and, therefore, contribute more to team activities. This sentiment is echoed in the final comment:

I do not think relying on students who are good at English leads to improvement. There is an inequality in student's English proficiency which results in differences in points. For example, in groups where there are both strong and weak English speakers, only the proficient individuals end up earning points.

The student makes a valid point, which could be one of the weaknesses of the game-based approach. It is possible that this effect might disillusion some higher-ability students. Moreover, less motivated students may just rely on more able members to earn points for the group. However, traditional non-game-based methods did not appeal to the majority of these students, and they overwhelmingly preferred the game-based approach. Thus, on balance, it demonstrated that it is an approach that has merit in this environment.

Conclusion

This study set out to incentivise students in mandatory English classes using gamified activities and to assess success based on students' attitudes toward this learning style. In terms of the research questions, the results provided an extremely promising view. Over 90% of students enjoyed the activities, felt motivated, and preferred them to conventional activities. However, there are lingering doubts in the literature about whether motivation is driven more by rewards rather than a desire to learn, and this concern was reflected in the results. Nevertheless, most of these students enjoyed the activities and felt motivated by this learning style. Compared to the start of the course, when most students expressed low motivation (through their dislike of English), the results must be viewed as a net positive. Even if some students were solely motivated by rewards, the evidence would seem to suggest that the learning outcomes of game-based learning are certainly no worse than traditional

methods. Therefore, the observed and perceived improvements in motivation and classroom atmosphere are valuable outcomes. Thus, this style of learning can be deemed suitable in an environment where students with low motivation are mandated to attend TOEIC classes.

One of the limitations of this study was that it relied on students' self-reporting. This was built into the study design, as the research question sought to investigate students' feelings or perceptions. As individuals can faithfully report their feelings, the findings should be accurate in that sense. However, the study does not provide evidence relating to objective measures of motivation that might more accurately predict outcomes. Another limitation was the size and homogeneity of the sample. All the participants were from a single institution. A larger sample taken from several universities would have made the results more generalisable.

Nonetheless, the student evaluations described in this paper support the existing literature, indicating that gamification can be an effective way to increase motivation. However, even when the introduction of game activities consistently causes increased student engagement to be observed, it does not necessarily translate directly to improved long-term outcomes (in excess of those gained through conventional teaching). Therefore, gamification can be viewed as one of the weapons within a teacher's armoury, ready to be deployed when students need incentivisation, rather than as a golden bullet that applies to every situation.

References

- Alomari, I., Al-Samarraie, H., & Yousef, R. (2019). The role of gamification technique in promoting student learning: A review and synthesis. *Journal of Information Technology Education: Research*, 18, 395–417. <https://doi.org/10.28945/4417>
- Antonaci, A., Klemke, R., & Specht, M. (2019). The effects of gamification in online learning environment: A systematic literature review. *Informatics*, 6(3), 1–22. <https://doi.org/10.3390/informatics6030032>
- Beirne, D. (2023). A classwork points system for grading classwork in university language courses. *Kobe JALT Journal*, 4(2), 4–17. https://www.kobejalt.org/_files/ugd/374330_418351f0b2914ff38fc05dee5cc1df78.pdf
- Buckley, P., & Doyle, E. (2014). Gamification and student motivation. *Interactive Learning Environments*, 24(6), 1162–1175. <https://doi.org/10.1080/10494820.2014.964263>
- Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, 125(6), 627–668. <https://doi.org/10.1037/0033-2909.125.6.627>
- Dehghanzadeh, H., Fardanesh, H., Hatami, J., Talaei, E., & Noroozi, O. (2019). Using gamification to support learning English as a second language: A systematic review. *Computer Assisted Language Learning*, 34(7), 934–957. <https://doi.org/10.1080/09588221.2019.1648298>
- Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: Defining “gamification”. *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments*, 9–15. <https://doi.org/10.1145/2181037.2181040>
- Dickinson, L. (1995). Autonomy and motivation: A literature review. *System*, 23(2), 165–174. [http://dx.doi.org/10.1016/0346-251X\(95\)00005-5](http://dx.doi.org/10.1016/0346-251X(95)00005-5)
- de Byl, P. (2013). Factors at play in tertiary curriculum gamification. *International Journal of Game-Based Learning*, 3(2), 1–21. <http://doi.org/10.4018/ijgbl.2013040101>
- Flores, J. F. F. (2015). Using gamification to enhance second language learning. *Digital Education Review*, 27, 32–54. <https://revistes.ub.edu/index.php/der/article/view/11912>
- Graham-Marr, A., Anderson, J., Howser, R., & Sato, A. (2016). *TOEIC Skills 1*. ABAX ELT Publishers
- Harmilawati, Sabaruddin, Kareba, A. M., Khaerunnisa, & Jusmaniar. (2023). Gamification in the learning evaluation on students' learning outcomes of English education students. *JLE: Journal of Literate of English Education Study Program*, 4(2), 61–70. <https://doi.org/10.47435/jle.v4i02.2273>
- Hanus, M. D., & Fox, J. (2015). Assessing the effects of gamification in the classroom: A longitudinal study on intrinsic motivation, social comparison, satisfaction, effort, and

- academic performance. *Computers & Education*, 80, 152–161.
<https://doi.org/10.1016/j.compedu.2014.08.019>
- Hung, A. C. Y. (2017). A critique and defense of gamification. *Journal of Interactive Online Learning*, 15(1), 57–72. <https://www.ncolr.org/jiol/issues/pdf/15.1.4.pdf>
- Laura-De La Cruz, K. M., Noa-Copaja, S. J., Turpo-Gebera, O., Montesinos-Valencia, C. C. Bazán-Velasquez, S. M., & Pérez-Postigo, G. S. (2023). Use of gamification in English learning in higher education: A systematic review. *Journal of Technology and Science Education*, 13(2), 480–497. <https://doi.org/10.3926/jotse.1740>
- Matsumoto, K. (1989). An analysis of a Japanese ESL learner's diary: Factors involved in the L2 learning process. *JALT Journal*, 11(2), 167–192.
<https://jalt-publications.org/sites/default/files/pdf-article/jj-11.2-art2.pdf>
- Mee, R. W. M., Shahdan, T. S. T., Ismail, M. R., Ghani, K. A., Pek, L. S., Von, W. Y., Woo, A., & Rao, Y. S. (2020). Role of gamification in classroom teaching: Pre-service teachers' view. *International Journal of Evaluation and Research in Education*, 9(3), 684–690. <http://doi.org/10.11591/ijere.v9i3.20622>
- Mekler, E. D., Brühlmann, F., Tuch, A. N., & Opwis, K. (2017). Towards understanding the effects of individual gamification elements on intrinsic motivation and performance. *Computers in Human Behavior*, 71, 525–534.
<https://doi.org/10.1016/j.chb.2015.08.048>
- Nurlely, L. (2024). Exploring gamification approaches in grammar learning. *FLIP: Foreign Language Instruction Probe*, 3(1), 67–75. <https://doi.org/10.54213/flip.v3i1.398>
- Pulfrey, C., Darnon, C. & Butera, F. (2013). Autonomy and task performance: Explaining the impact of grades on intrinsic motivation. *Journal of Educational Psychology*, 105(1), 39–57. <https://doi.org/10.1037/a0029376>
- Rahmani, E. (2020). The benefits of gamification in the English learning context. *IJEE (Indonesian Journal of English Education)*, 7(1), 32–47.
<http://dx.doi.org/10.15408/ijee.v7i1.17054>
- Tamayo, M. R., Cajas, D. & Sotomayor, D. D. (2023). Using gamification to develop vocabulary and grammar among A1 level of English students: A quasi-experimental design. In: M. Botto-Tobar, M. Zambrano Vizuete, S. Montes León, P. Torres-Carrión, & B. Durakovic, B. (Eds), *Applied Technologies (ICAT 2022)*. Communications in Computer and Information Science, vol. 1757. Springer, Cham.
https://doi.org/10.1007/978-3-031-24978-5_16
- Werbach, K., & Hunter, D. (2012). *For the win: How game thinking can revolutionise your business*. Wharton School Press.
- Yaccob, N. S., Rahman, S. F. A., Mohamad, S. N. A., Rahim, A. A. A., Rashid, K. K. A., Aldaba, A. M. A., Yunus, M. M., & Hashim, H. (2022). Gamifying ESL classrooms through gamified teaching and learning. *Arab World English Journal (AWEJ) Special Issue on CALL*, 8, 177–191. <https://dx.doi.org/10.24093/awej/call8.12>

Appendix

Questionnaire

1. Please rate how strongly you enjoy playing games in class activities.
 1. I really dislike it
 2. I dislike it a little
 3. Sometimes I like it, sometimes I don't like it
 4. I like it a little
 5. I really like it
2. Do you feel you learn more in a game-based activity or more in an activity that is non-game-based (worksheet, or interview for example)?
 - a. Game-based learning
 - b. Non-game-based learning
3. Do you feel more motivated by a game-based activity or more motivated by an activity that is non-game-based (worksheet, or interview for example)?
 - a. Game-based learning
 - b. Non-game-based learning
4. When you play a game in class, are you focused on getting the points or language learning?
 - a. Points
 - b. Language learning
 - c. Both
5. Has the use of games and teamwork made TOEIC classes more enjoyable?
 - a. Yes
 - b. No
6. Do you believe you have learned useful language and communication skills through the games and teamwork used in TOEIC classes?
 - a. Yes
 - b. No
7. Do you believe your TOEIC skills have improved as a result of the games and teamwork used in classes?
 - a. Yes
 - b. No
8. Has your TOEIC score improved over the past year?

a. Yes

b. No

9. What is your opinion about using games in the classroom? Please write a few sentences about your views.

1 クラス活動の中でのゲームにどのくらい楽しめたかを評価してください。

1 とても楽しめていない

2 少し楽しめなかった

3 楽しめている時もあり、時折楽しめていない時もある

4 少し楽しめている

5 とても楽しめている

2 ゲーム重視のクラス活動でより学びを得ることが出来ますか、それともゲーム以外の活動の方がそうだと感じますか？(例えば、ワークシートやインタビューなどの活動)

a ゲーム重視での学び

b ゲーム重視では無い学び

3 ゲーム重視の活動に対してモチベーションを感じますか、それともゲーム以外の活動の方がモチベーションになりますか？

a ゲーム重視の学び

b ゲーム重視では無い学び

4 クラスでゲームをしている際、ポイントを取ることに集中していますか、それとも言語を学ぶということに集中していますか？

a ポイントを取ることに

b 言語を学ぶことに

c どちらも

5 ゲームやチームワークを通して、TOEIC のクラスはもっと楽しいものになりますか？

6 TOEIC のクラスで、ゲームやチームワークを通して有用な言語やコミュニケーション能力を学べたと思いますか？

7 クラス内でのゲームやチームワークを通してあなたの TOEIC スキルが良くなったと思いますか？

8 これまでと比べ、あなたの TOEIC の結果は良くなりましたか？

9 クラス内でゲーム活動を用いることに対してどう思いますか？あなたの意見や考えを文章でご回答ください。

Effects of COVID-19 Learning Loss on Freshman English Test Scores

Mark Goodhew, Asia University

Abstract

Learning Loss became a common term during the COVID-19 pandemic to describe the decline in academic achievement witnessed internationally in some scholastic populations. One place this effect was noticed was in the results of standardized tests, but almost all the focus was on pre-tertiary K–12 education. Less data is available for the tertiary level, and international results have been mixed. The author, evaluating Freshman English students at a Tokyo university, found no statistically significant difference in test scores between the 2021–22 remote/hybrid teaching year and the following 2022–23 and 2023–24 in-person teaching years ($n = 166$). A broader topic of inquiry, however, would be what potential for learning was lost or hindered due to the two uniquely challenging years of remote/hybrid teaching.

Introduction

Learning Loss (LL) can be defined as “any specific or general loss of knowledge and skills or to reversals in academic progress, most commonly due to extended gaps or discontinuities in a student’s education” (The Glossary of Education Reform, 2013). LL can be caused by various factors and take a variety of forms. Prior to the COVID-19 pandemic, the most discussed and researched form was Summer Learning Loss, the LL that occurs during the extended summer holidays when some countries’ schools break for over two months. This is a well-documented and researched problem (Alexander et al., 2016; Cooper et al., 1996), and several educational reforms have been proposed to mitigate its effects (Borman & Boulay, 2004; Cooper, 2003). A common strategy is the implementation of Expanded Learning Time, where school days and academic calendars are reconfigured to increase the total amount of time spent in school per year, and often to reduce the length of these long holidays (Rocha, 2007).

LL took on new significance and became better known during the COVID-19 pandemic when schools and universities worldwide were suddenly forced to close their doors. Some were able to transition to remote learning, while others were not. The focus turned to how well students were adapting to these pandemic mitigation policies and to what degree they were falling behind in their education.

Within the context of this larger conversation on LL, certain questions should be considered: What LL occurred at the tertiary level according to test scores and other assessment data? How did LL affect past students—those who took classes using remote and hybrid teaching modalities—compared to current students who are once again taught in person? This paper will first review research on LL at the tertiary level internationally. Then it will report the test results of the author’s Freshman English students over the 2021–22, 2022–23, and 2023–24 academic years, comparing these distinct “Zoom” and “Post-Zoom” eras. Lastly, the findings and implications of this analysis will be discussed.

Literature Review

Most, if not all, of the conversation on LL has focused on pre-tertiary K–12 education, as this is the scholastic cohort on which governments and families primarily focus. In addition, from a research perspective, much more data is available on this cohort. Many countries require students at specific grades to take standardized tests, which are then used to track how well schools and school districts are doing and to compare results nationally and

internationally. Perhaps the most well-known of these international comparisons is the OECD's Programme for International Student Assessment (PISA), which tests 15-year-old students in math, reading, and science (OECD, 2024).

No national or international assessments exist for students currently attending university. For most students, the last time they take a standardized test is when leaving secondary school or applying to university. Thus, similar comparisons cannot be made of university students of different years and countries, and the literature on tertiary COVID-19 LL that includes assessment data is limited. Instead, much of the literature is on the mental health, motivation, and well-being of university students during the pandemic, when the teaching modalities changed from in-person to remote.

According to the literature that does focus on or include assessment, the international results are mixed. One study compared the attitude of Japanese medical students towards e-learning in relation to their performance on computer-based achievement tests. The 2020 and 2021 scores were, in fact, considerably higher than those of the previous three years, possibly because students had more time to study at home due to the pandemic. Interestingly, while more than half of the students preferred asynchronous on-demand classes (54.2%), students who preferred face-to-face classes (19.9%) had much higher scores than their peers (Sekine et al., 2022). In another study of 458 students across three majors at a Spanish university, Gonzalez et al. (2020) reported "a significant positive effect of the COVID-19 confinement on students' performance." They found that students did not study continuously before the pandemic and that the COVID-19 confinement changed their study habits into a more continuous routine, improving their performance.

By contrast, Orlov et al. (2021), in a study of economics students in seven courses across four universities in the USA, found that average total assessment scores declined by 0.2 standard deviations. However, there were substantial variations in outcomes across courses: prior online teaching experience by the instructor and the use of teaching methods that encouraged active engagement strongly mitigated this adverse effect. In another study, Motz et al. (2021) surveyed 6,156 undergraduate students at Indiana University and analyzed their Canvas LMS data. They found that when the university transitioned to remote instruction, there was an increase in the number of assignments that students had to complete. Interestingly, students who spent more time and reported more effort on these assignments generally had lower course scores. They argue that in the process of rapidly transitioning to remote learning, instructors increased the number of "busywork" assignments, which had a

negative effect on student outcomes. Excessive assignments led to less effort per assignment and lower overall course scores.

Permana et al. (2023) surveyed 100 university students across the island of Java in Indonesia. They found that while most students reported health issues such as eyesight problems and exhaustion, experienced a decrease in learning motivation, and had trouble concentrating, most self-reported their academic scores as unchanged (65%), with 28% reporting an increase and 7% reporting a decrease, as compared to their pre-COVID-19 results. In contrast, Tan (2021), in a survey of 282 university students across Malaysia, found that students self-reported a decrease in motivation and learning performance, with primary causes being a lack of necessary infrastructure to support learning and a loss of social support from lecturers and peers.

Two studies on Japanese university students' TOEIC scores have also reported mixed results. Nagata (2022) studied Reitaku University students who used an online English learning system, which combined one-on-one remote English conversation and a separate English training software, for the 2020–21 academic year. He found that average TOEIC Listening scores increased by 0.52 per hour and Reading scores by 0.35, which is twice the increase for Listening and 1.7 times for Reading when compared to the university's regular English classes. However, there was a large variance in the usage of this online system among students due to varying motivation. In contrast, Richard (2023) studied four cohorts of students at a prefectural university who took the TOEIC test at the beginning and end of their two-year English program. While all four cohorts made gains, the cohort that experienced two years of remote learning made fewer gains in Listening than the other three cohorts and fewer gains in Reading than the subsequent cohort. The two years of remote/hybrid teaching were thus found to have had a negative effect on TOEIC scores, though Richard has some doubts about these conclusions, as the quality of the online TOEIC test when it was introduced in 2020 is questioned.

Within this context of mixed results on the academic impact of remote/hybrid teaching during the pandemic, the following research questions are addressed in this three-year study of Japanese university students:

RQ1. How did the test scores of the author's students compare by class and year?

RQ2. Did the remote/hybrid teaching have a statistically significant effect on these test scores?

Methodology

Context

The author joined the university at the start of the 2021–22 academic year. Due to COVID-19, the previous 2020–21 academic year had been conducted entirely online via synchronous Zoom lessons. As encouraged by the Ministry of Education, Culture, Sports, Science and Technology (MEXT), the 2021–22 year began as a hybrid system: students could attend lessons in-person or stay at home and participate via Zoom. However, Tokyo prefecture entered a state of emergency during Week 1 of the Spring semester, and the university had to revert to remote teaching from Week 2 through Week 9. The last four weeks of the term, Week 10 through Week 13, were conducted with the hybrid system, as was the entire following Fall semester. Throughout this year, the vast majority of students chose to attend lessons remotely, and it was not uncommon for a class of 20 students to have only one or two attend in-person. From the following 2022–23 academic year, Zoom was not used, and all classes were conducted in-person.

Participants and Tests

Freshman English (FE) is a two-semester course, and classes are held multiple times a week. Students are leveled into groups of similar English ability based on their TOEIC IP scores, which they take prior to the beginning of the year. Instructors are expected to balance their lessons between an assigned four-skills textbook and a TOEIC textbook, the levels of which differ depending on the level of students to be taught. Unit tests provided by the publishers of both textbooks were converted to Google Form tests by the author and administered to his three FE classes (Class A, Class B, Class C) per year for the 2021–22, 2022–23, and 2023–24 academic years (nine classes over three years, $n = 166$). Five four-skills tests and four TOEIC tests were administered to Class A and Class B, who were seen four days a week, and three four-skills tests and two TOEIC tests were administered to Class C, who were seen four days a week in the Spring semester but only twice a week in the Fall semester. The same tests were administered in all three years. In 2021–22 most students took the tests remotely from home, and in 2022–23 and 2023–24 students took the tests in the classroom using their laptops or other personal devices.

Analyses

For RQ1, descriptive statistics were used.

For RQ2, a one-way ANOVA and a linear mixed model (LMM) analysis were conducted.

Results

RQ1. How did the test scores of the author's students compare by class and year?

The analysis revealed that, from 2021–24, Class A's four-skills tests (Class A FS) had a mean score of 82.60% (SD = 7.67%), with scores spanning from 57.58% to 100.00%. In contrast, their TOEIC tests (Class A T) showed a mean score of 52.21% (SD = 10.36%), ranging from 25.00% to 76.79%. Class B FS had an average score of 80.00% (SD = 8.84%), with a range of 41.67% to 96.92%. The mean score for Class B T was 51.53% (SD = 10.72%), ranging from 25.00% to 80.36%. For Class C FS, scores averaged 78.16% (SD = 12.17%), ranging from 34.85% to 98.18%. Lastly, Class C T had an average of 53.75% (SD = 11.76%), with scores between 23.21% and 73.21%. These contrasting results indicate that students generally performed better on their four-skills tests compared to their TOEIC tests. Overall, across all classes and tests, the mean score was 68.25% (SD = 17.30%), with scores ranging from 23.21% to 100.00%.

Table 1. *Scores by Class, 2021–24*

Class	Mean	SD	Min	Max
Class A FS	82.60%	7.67%	57.58%	100.00%
Class A T	52.21%	10.36%	25.00%	76.79%
Class B FS	80.00%	8.84%	41.67%	96.92%
Class B T	51.53%	10.72%	25.00%	80.36%
Class C FS	78.16%	12.17%	34.85%	98.18%
Class C T	53.75%	11.76%	23.21%	73.21%

When analyzing scores by year, 2021–22 had a mean score of 68.30% (SD = 17.13%), with scores ranging from 32.14% to 98.18%. 2022–23 showed a mean of 68.33% (SD = 17.24%), ranging from 25.00% to 96.92%. For 2023–24, the mean score was 68.13% (SD = 17.55%), with scores ranging from 23.21% to 100.00%. Overall, across all years, the mean score was 68.25% (SD = 17.30%), with scores spanning from 23.21% to 100.00%.

Table 2. Scores by Year

Year	Mean	SD	Min	Max
2021–22	68.30%	17.13%	32.14%	98.18%
2022–23	68.33%	17.24%	25.00%	96.92%
2023–24	68.13%	17.55%	23.21%	100.00%

RQ2. Did the remote/hybrid teaching have a statistically significant effect on these test scores?

The year with the highest mean score was 2022–23, with a mean of 68.33% (SD = 17.24%), 95% CI [66.55, 70.10]. The scores for this year ranged from 25.00% to 96.92%. Following very closely, the year 2021–22 had a mean score of 68.30% (SD = 17.13%), 95% CI [66.62, 69.97], with scores ranging from 32.14% to 98.18%. The year 2023–24 had the lowest mean score of 68.13% (SD = 17.55%), 95% CI [66.40, 69.85], with scores ranging from 23.21% to 100.00%. A one-way ANOVA was performed to compare the mean scores across the three years (2021–22, 2022–23, and 2023–24). The analysis showed no statistically significant differences between the mean scores for the years, $F(2, 1167) = 0.015, p = .985$.

Table 3. ANOVA Results

Year	Mean	SD	SE	95% CI for Mean		Min	Max
				LL	ULL		
2021–22	68.30%	17.13%	0.85%	66.62%	69.97%	32.14%	98.18%
2022–23	68.33%	17.24%	0.90%	66.55%	70.10%	25.00%	96.92%
2023–24	68.13%	17.55%	0.88%	66.40%	69.85%	23.21%	100.00%

A linear mixed model (LMM) analysis was conducted to examine the effects of class, test, and year on scores. The analysis included all 1170 test scores, with no exclusions. The categorical variables were distributed as follows: Class (Class A FS: 22.9%, Class A T: 17.9%, Class B FS: 20.2%, Class B T: 16.3%, Class C FS: 13.2%, Class C T: 9.4%), Test (FS Test 1: 13.7%, FS Test 2: 13.5%, FS Test 3: 13.1%, FS Test 4: 8.2%, FS Test 5: 7.9%, TOEIC Test 1: 13.7%, TOEIC Test 2: 13.4%, TOEIC Test 3: 8.5%, TOEIC Test 4: 8.0%), and Year (2021–22: 34.6%, 2022–23: 31.2%, 2023–24: 34.2%). The dependent variable, score, had a mean of 68.25% (SD = 17.30), ranging from 23.21% to 100.00%.

Goodness-of-fit statistics indicated that the model fit the data adequately (Deviance/df = 299.636, Pearson Chi-Square/df = 299.636, AIC = 9997.319). The likelihood ratio chi-square test comparing the fitted model against the intercept-only model was not significant, $\chi^2(2) = 0.030, p = .985$, indicating that year did not significantly improve the model.

The intercept was significant, $B = 68.129, SE = 0.864, 95\% CI [66.435, 69.824], p < .001$, reflecting the overall mean score. However, the effect of year on scores was not significant: 2021–22 ($B = 0.167, SE = 1.219, 95\% CI [-2.221, 2.556], p = .891$) and 2022–23 ($B = 0.198, SE = 1.251, 95\% CI [-2.254, 2.651], p = .874$), with 2023–24 serving as the reference year. Overall, the analysis revealed significant differences in scores between different classes and tests, but no significant differences were found across the years.

Table 4. *Linear Mixed Model Results*

Variable	Category/Statistic	Percentage/Value
Goodness-of-Fit Statistics		
Deviance/df		299.636
Pearson Chi-Square/df		299.636
AIC		9997.319
Model Comparison	Likelihood Ratio Chi-Square Test	$\chi^2(2) = 0.030, p = .985$
Fixed Effects		
Intercept	$B = 68.129$	$SE = 0.864, 95\% CI [66.435, 69.824], p < .001$
Year (2021–22)	$B = 0.167$	$SE = 1.219, 95\% CI [-2.221, 2.556], p = .891$
Year (2022–23)	$B = 0.198$	$SE = 1.251, 95\% CI [-2.254, 2.651], p = .874$
Year (2023–24)	$B = 0.000$	$SE = 1.000, 95\% CI [-2.00, 2.000], p = 1.000$

Discussion

For the scholastic cohort of the author's FE students, no statistically significant effect on test scores was found between the 2021–22 remote/hybrid teaching year and the following 2022–23 and 2023–24 in-person teaching years. This finding is congruent with the previous literature on tertiary LL, surveyed above, which reported mixed results. Gonzalez et al. (2020), Nagata (2022), and Sekine et al. (2022) found that performance had improved, while Motz et al. (2021), Orlov et al. (2021), Richard (2023), and Tan (2021) found that performance had declined. The results of this study are consistent with those of Permana et al. (2023), who found that academic performance was primarily unchanged. A drawback of all these tertiary-level studies is that they are limited in scope and relevant only to the specific

cohort studied, and do not represent entire age groups in the way that standardized national and international assessments do. A broader analysis of what LL occurred generally at the tertiary level cannot be explored through these discrete results.

This specific study of the author's FE students has its limitations. The data begins in the middle of the remote/hybrid teaching period, as that is when the author joined the university. A more extensive data series would strengthen the analysis—for example, if there were data on test scores prior to COVID-19 and the entire two years of the remote/hybrid era. Another limitation is that this study only includes students' FE test scores, and not their overall university test scores or GPA. Nevertheless, the fact that such an immense shock as remote/hybrid teaching during COVID-19 had no noteworthy effect on these test scores when compared to the following two post-COVID-19 in-person teaching years is in itself notable.

In contrast, one study on the post-COVID-19 PISA tests found that school closures led to significant learning loss, with scores declining on average by 14% of a standard deviation or roughly seven months of learning. Losses were notably greater in schools that stayed closed for longer (Jakubowski et al., 2024). However, as discussed in the literature review, what occurred at the pre-tertiary level cannot be equated with the tertiary level. While both the PISA tests and this study concern themselves with academic performance, the results of skills tests taken by 15-year-old students should perhaps not be likened to the EFL test results of Japanese university students.

Some differing features between the first year and the following two years of the study should also be mentioned. Compared to in-person instruction, remote instruction made it much easier for students to lose focus and passively observe the lesson. On the positive side, students likely had more time to study, as Sekine et al. (2022) and Gonzalez et al. (2020) noted in their studies. By contrast, when students take lessons in-person, they are usually much more engaged with the material at that moment, as they must accomplish various activities as directed and supervised by the instructor. Due to busier schedules, however, they may be spending less time studying.

Conclusion

It should be noted that there has been pushback by some against this focus on LL. One claim is that the test score declines seen in national assessments were relatively modest, given how devastating the pandemic was and how it upended nearly every aspect of daily life at the time (Wallace-Wells, 2022). Another argument is that the concept of COVID-19 LL

should itself be questioned because it is a version of deficit thinking that creates unneeded anxiety for students, parents, and teachers. LL does not acknowledge all the hardships that were endured, and all the things that were learned, during the pandemic outside the confines of academic school studies (City Year, 2021).

This study analyzes the topic of COVID-19 LL as it relates to test scores and assessment data, intentionally limiting the scope of the analysis to this domain. However, LL as a concept is much more than its effect on test scores, in the same way that learning and education can never be fully explicated by assessment. Future studies may want to include qualitative or mixed-methods approaches to provide a more nuanced understanding. For the scholastic cohort of Japanese university students, a broader topic of inquiry would be what potential for learning was lost or hindered due to the two uniquely challenging years of remote/hybrid teaching.

References

- Alexander, K., Pitcock, S., & Boulay, M. C. (2016). *The summer slide: What we know and can do about summer learning loss*. Teachers College Press.
- Borman, G. D. & Boulay, M. (2004). *Summer learning: Research, policies, and programs*. Routledge.
- City Year. (2021, June 15). *Why we don't use the term "learning loss"*.
<https://www.cityyear.org/national/stories/education/why-we-dont-use-the-term-learning-loss/>
- Cooper, H., Nye, B., Charlton, K., Lindsay, J. O., & Greathouse, S. (1996). The effects of summer vacation on achievement test scores: A narrative and meta-analytic review. *Review of Educational Research*, 66(3), 227–268.
<https://doi.org/10.3102/00346543066003227>
- Cooper, H. (2003). Summer learning loss: The problem and some solutions. *ERIC Digest*.
<http://files.eric.ed.gov/fulltext/ED475391.pdf>
- The Glossary of Education Reform. (2013, August 29). *Learning loss definition*.
<https://www.edglossary.org/learning-loss/>
- Gonzalez, T., De La Rubia, M., Hincz, K. P., Comas-Lopez, M., Subirats, L., Fort, S., & Sacha, G. M. (2020). Influence of COVID-19 confinement on students' performance in higher education. *PLOS ONE*, 15(10), e0239490.
<https://doi.org/10.1371/journal.pone.0239490>
- Jakubowski, M., Gajderowicz, T., & Patrinos, H. A. (2024). *COVID-19, school closures, and student learning outcomes: New global evidence from PISA* (EdWorkingPaper No. 24-899). Annenberg Institute at Brown University. <https://doi.org/10.26300/f8p1-ms45>
- Motz, B. A., Quick, J. D., Wernert, J. A., & Miles, T. A. (2021). A pandemic of busywork: Increased online coursework following the transition to remote instruction is associated with reduced academic achievement. *Online Learning Journal*, 25(1), 70–85. <http://doi.org/10.24059/olj.v25i1.2475>
- Nagata, M. (2022). Quantitative analysis of the effects of one-on-one online English conversation learning on TOEIC scores. *Reitaku Journal of Economic Studies*, 29(1), 12–23.
- OECD. (2024). *Programme for international student assessment (PISA)*.
<https://www.oecd.org/en/about/programmes/pisa.html>
- Orlov, G., McKee, D. E., Berry, J. D., Boyle, A., DiCiccio, T. J., Ransom, T., Rees-Jones, A., & Stoye, J. (2021). Learning during the COVID-19 pandemic: It is not who you teach, but how you teach. *Economics Letters*, 202.
<https://doi.org/10.1016/j.econlet.2021.109812>

- Permana, A. A., Kusnadi, A., Marpaung, A. N., Marcela, E. D., Yuasan, N., & Fauziyah, S. (2023). Effect of Zoom fatigue on health and learning loss in students during the COVID-19 pandemic. *Jurnal Inovasi Teknologi Pendidikan*, 10(1), 22–32. <https://doi.org/10.21831/jitp.v10i1.52930>
- Richard, J. J. (2023). Using difference-in-differences to compare cohort-level TOEIC L&R scores. *Shiken: A Journal of Language Testing and Evaluation in Japan*, 27(1), 1–11. <https://doi.org/10.37546/JALTSIG.TEVAL27.1-1>
- Rocha, E. (2007). Choosing more time for students: The what, why, and how of expanded learning. *Center for American Progress*. https://cdn.americanprogress.org/wp-content/uploads/issues/2007/08/pdf/expanded_learning.pdf
- Sekine, M., Watanabe, M., Nojiri, S., Suzuki, T., Nishizaki, Y., Tomiki, Y., & Okada, T. (2022). Effects of COVID-19 on Japanese medical students' knowledge and attitudes toward e-learning in relation to performance on achievement tests. *PLOS ONE*, 17(3), e0265356. <https://doi.org/10.1371/journal.pone.0265356>
- Tan, C. (2021). The impact of COVID-19 on student motivation, community of inquiry and learning performance. *Asian Education and Development Studies*, 10(2), 308–321. <https://doi.org/10.1108/AEDS-05-2020-0084>
- Wallace-Wells, D. (2022, October 30). Opinion | Pandemic learning loss is not an emergency. *The New York Times*. <https://www.nytimes.com/2022/10/29/opinion/test-scores-covid-pandemic-learning-loss.html>

Japanese University Students' Perspective on NEST & N-NEST

Armando Moreno, Asia University

Joshua Trued, Asia University

Abstract

This article examines 224 first-year Japanese university students' attitudes and perceptions toward Native English-Speaking Teachers (NEST) and Non-Native English-Speaking Teachers (N-NEST) using a mixed-methods approach. A Likert scale survey and an open-ended survey were administered to gain deeper insights into the students' preferences and experiences. The quantitative results indicated that students may have less hesitation to ask N-NEST questions compared with their NEST counterparts. The qualitative responses revealed that regardless of an instructor's status as a native English speaker, students focus on personality-related features in determining their ideal teacher. Furthermore, the results showed that students are unable to ascertain whether or not their teachers are NESTs or N-NESTs, and that while NESTs are valued for their pronunciation, N-NESTs are valued for their approachability.

Introduction

English is widely recognized as a global language for cross-cultural communication in numerous nations, and it should come as no surprise that an increasing number of English language teachers would be considered as “non-native” speakers. It has been reported that nearly 80 percent of English teachers would fall under the category of a Non-Native English Speaking Teacher, or N-NEST (Selvi, 2014). This is due to the fact that the population of English language learners has surpassed that of native English-speaking teachers. The increasing requirements of English learners necessitate a rising number of N-NESTs due to the insufficient number of available Native English Speaking Teachers, or NESTs.

Perhaps due to this trend, research interest concerning non-native English-speaking instructors has experienced significant growth. One might initially perceive non-native English-speaking instructors as deficient due to their fluency and cultural background. If linguistic competence were the exclusive measure of teaching effectiveness, it would follow that native English-speaking teachers would be inherently more esteemed. However, this line of thinking may be flawed; indeed, according to students’ perception, N-NESTs have a greater capacity to act as role models for proficient language learners, may exhibit a higher level of empathy toward the challenges faced by their students as they themselves have studied English as an L2, and may have a greater ability to foresee potential challenges in language acquisition (Medgyes, 1992).

The study reported in this paper attempts to answer the following research questions:

1. How do Japanese students perceive their lessons being taught by a NEST compared to a N-NEST in a university context?
2. Can students perceive a difference between NESTs and N-NESTs?
3. Regardless of the N-NEST/NEST distinction, what qualities do university students seek in an instructor?

Literature Review

NESTs vs. N-NESTs

EFL students’ preference regarding native-speaking English instructors versus non-native English-speaking instructors varies based on different studies. Mahboob (2004) demonstrates that students perceive the speaking proficiency, lexical competence, and cultural awareness as notable attributes of NESTs. Conversely, according to other research,

NESTs regard their role as cultural advisors to be of lesser significance compared to their role as language educators who prioritize enhancing students' speaking abilities (Walker, 2001). In an ESL context, students in American and British English Language Programs (ELPs) preferred to be instructed by NESTs (Mahboob, 2005). In two different studies, students who were enrolled in English programs in the US were found to hold positive opinions of grammar classes instructed by NESTs yet were not shown to hold negative opinions toward N-NESTs in terms of listening and speaking (Moussu, 2006). In Thailand, 261 English learners preferred NESTs even though they held N-NESTs in high esteem (Todd & Pojanapunya, 2009).

Another study showed that students who prefer NESTs tend to prioritize areas such as vocabulary, pronunciation, culture, attitudes, and assessment methods. In contrast, for listening, grammar, reading, and strategies related to language learning, students leaned towards N-NESTs. Significant differences were found in preferences for reading based on experience with NESTs. Overall, there was a general preference for NESTs, but the preference was higher for a combination of both NESTs and N-NESTs (Lasagabaster & Sierra, 2002).

In a different study, 50 Japanese students studying English at a university in Japan cited pronunciation as the biggest disadvantage of N-NESTs. An issue frequently raised pertained to the matter of precision, as indicated by the claim that there are N-NESTs who lack the ability to accurately articulate words (Walkinshaw & Oanh, 2014). However, it should be noted that the same study found that students perceived communication with N-NESTs to be easier as a result of their shared cultural background, a phenomenon also observed by Ling and Braine (2007). One student commented that they were able to ask their N-NEST teacher a favor without hesitation, while approaching their NEST was a cause of stress for them.

Research in Saudi Arabia found that while being a native speaker was not crucial, students emphasized the importance of pedagogical skills and teaching abilities (Murtada, 2023). Another study indicates that educators who do not fit the stereotype of native English-speaking teachers encounter difficulties when engaging with students and suggested promoting intercultural awareness as a successful educational tactic in helping students cultivate a mindset of tolerance and acceptance towards individuals from diverse ethnic and cultural backgrounds (Egitim & Garcia, 2021).

Research has not wholly found that students perceive NESTs to be superior to N-NESTs. A study of 420 students enrolled at a university in Hong Kong demonstrated a predominantly positive disposition towards N-NESTs whose perceived effectiveness was on par with that of native-speaker teachers (Ling and Braine, 2007). A study conducted in Iran revealed that students of NESTs showed greater fluency and lexical complexity, while students taught by N-NESTs showed higher accuracy in spoken language production (Ghane & Razmi, 2023). In a study done in Taiwan, 45% of students stated that due to having a shared L1 and cultural background, learning with a N-NEST was preferable (Tsou & Chen, 2017). (Here, the term “N-NEST” is used to include local teachers who share an L1 with students but does not include information about N-NEST teachers who may not share an L1 with students, a common trend in the literature.) Despite findings such as these that positively portray them, N-NEST teachers have voiced the concern that they are seen as lacking (Wang, 2012).

A different study done on the perceptions of both NEST and N-NEST held by Armenian graduate students found that they were both thought to be effective. These groups had a preference for NESTs’ pronunciation, while for N-NESTs, their vocabulary, emotional support, and cultural understanding were thought to be superior. In general, students did not present a strong bias towards N-NESTs, as they concluded that both groups were capable of producing effective instruction that matched the students’ expectations. On the other hand, the study revealed that students had significant confusion regarding the term “native speaker,” which implies that the participants didn’t understand the ideologies related to native-speakerism and its implications in language teaching. Overall, the Armenian students valued positive teaching attitudes such as enthusiasm, care, and effective pedagogical skills more than the teacher’s native-speaker status (Lilit, 2024).

A Japanese Context

Traditionally, in Japanese secondary school, English is taught by a Japanese teacher in the Japanese language in a classroom of around 30 students. Foreign teachers were introduced into this classroom environment in 1987 when the Japan Exchange and Teaching (JET) program was created with the objective to focus more on students’ oral communication in English. (Tajino & Tajino, 2000). The reason why foreign educators were seen as necessary was because Japanese educators prioritize the instruction of grammar and vocabulary, whereas educators from other countries often involve students in oral

communication exercises that emphasize student participation (Esaki & Shintani, 2010). The co-teaching model is commonly experienced by most public school students in Japan until they transition to university. Consequently, certain students might form perceptions about foreign English teachers prior to even starting university (Johannes, 2012). These preexisting beliefs could lead to the formation of teacher stereotypes and influence the dynamics between teachers and students.

Studies have shown the possibility of specific societal notions regarding the ideal foreign English teacher in Japan (Nagatomo, 2016). For example, much of the research regarding the NEST/N-NEST dichotomy includes cases where Japanese-speaking English instructors are put into the N-NEST category. At the university where the current study was conducted, the N-NEST group of teachers did not include any Japanese instructors. However, it should be mentioned that some students meet with a Japanese teacher of English once per week for 50 minutes. (They meet with a NEST four times per week for 50 minutes.) This is important to note because a study by Yazawa (2017) showed that a majority of 320 high school students were more partial to their Japanese teacher of English over their foreign teacher, mainly because of a shared L1, a feature not shared by the N-NEST group in this study. However, about one third of the students in the study preferred learning with their foreign English teacher because of “authentic” English and pronunciation.

Research conducted by Cumming and Matsumoto (2018) revealed distinctions among high school and university students concerning diverse teacher-related aspects including classroom conduct, character traits, and pedagogical abilities. University students perceive their English teachers as more influential on their motivation compared to high school students. This suggests that university students place higher importance on these aspects of their teachers in influencing their motivation levels, which might imply that students at a tertiary level see their teacher as a stronger influence and someone who might directly impact their interest in the subject and their development and performance. A similar study noted that Japanese students' perceptions regarding their teachers were investigated as a possible motivational factor. The investigation also analyzed the teacher-related aspects (including personality, teaching-related skills, classroom behavior, and other variables) that students identified as having the greatest impact on their motivation.

Japanese university students have varied perceptions of N-NESTs from countries other than Japan. Some students prefer to have English instructors of American/European heritage, believing that they can teach the language and culture more effectively. There are

also students who hold stereotypes about N-NESTs, particularly those of Asian descent, and question their ability to teach effectively (Egitim & Garcia, 2021). In addition, students associate English-speaking abilities with being American and often assume that native English speakers are white individuals, not necessarily Africans or Asians. These perceptions are influenced by factors such as gender and personality stereotypes as well as mass media, education, and culture.

Significant differences were found in Japanese students' views on teachers between NESTs and N-NESTs (Matsumoto, 2017). Students indicated that NESTs encouraged a higher level of enthusiasm on their motivation to learn. This disparity demonstrated statistical significance, suggesting that Japanese EFL students at high school and university levels may view NESTs as more influential in motivating them to learn English compared to N-NESTs. Nevertheless, additional research has indicated that distinct assessments and underlying connections pertaining to educators, alongside the implicit favoritism exhibited by students towards their own social group, exhibit no notable variations depending on the teacher's native speaker status (Oude & Müller, 2024).

Research Method

Setting and Participants

The participants for this study come from a Japanese university located in Tokyo. The university's main faculties are Business, Economics, Law, International Relations, Urban Innovation, Data Science, and Multicultural Communications. The university also has study-abroad programs and support for international students. There are 28 full time English instructors at the university comprising both N-NESTs and NESTs. Requirements for employment include a master's degree related to TESOL, EFL, or ESL as well as relevant teaching experience. Importantly, native speaker status is not a requirement for employment. A total of 224 students were interviewed, with the majority (212 students) coming from mainly a Japanese cultural background. Their ages range from 18 to 25. Out of the interviewed students, 84 were female subjects and 140 were male subjects. Most of the Japanese students have not experienced living in a country other than Japan, and only a handful have traveled abroad. The students participated voluntarily and were not rewarded monetarily for their responses.

Data Collection Instrument

The data collection tool employed for this study was an anonymous self-report Likert-scale questionnaire in both English and Japanese, followed by a section to collect short answers. This questionnaire aimed to capture the attitudes of learners in relation to their engagement with the English language, specifically in connection to various types of instructors. Teachers were sent information and a link to Google Forms and were asked to send said link to students as well as encourage them to complete it. The data collection was performed during the autumn of 2023. Two hundred and twenty-four complete responses were analyzed and sorted into NEST and N-NEST categories, with the NEST category including 57 students and the N-NEST category including 124 students. This facilitated data collection in a format conducive to efficient analysis. Anonymity was emphasized in this study, so students' personal identities have not been revealed. Its methodological design drew parallels with previous investigations centered on similar themes and reliant upon questionnaire-based methodologies (Benke & Medgyes, 2005; Chiba et al., 1995; Walkinshaw & Duong, 2012).

The questionnaire starts by asking about the general background of the students in the form of simple questions regarding age, gender, cultural and national background, and previous experiences learning English. Finally, students were asked if their current teacher was a native or a non-native speaker.

After these initial questions, students were presented with a series of statements and asked to rank how much they agree or disagree with them on a scale of 1 to 5, where 1 means "strongly disagree" and 5 means "strongly agree." The questions for this section were aimed at assessing various aspects of an English teacher's effectiveness and the overall learning experience in an English class. They cover a range of factors such as the teacher's ability to explain grammar, pronunciation, accessibility, fostering a positive learning environment, cultural knowledge, exposure to colloquial language, understanding of literature and idiomatic expressions, and the impact of the teacher's nationality on the learning experience. Students were also asked if they would recommend either NEST or/and N-NEST to other English language learners.

The final set of questions are designed for the students to write their opinions in either English or Japanese. In fact, to facilitate the elicitation of comprehensive and holistic viewpoints on the matter at hand, an open-ended format was employed in the questionnaire's design (see Table 2).

In sum, this questionnaire directly addresses whether students prefer to learn from a native or non-native English teacher, with the aim of understanding individual preferences and perceptions towards language teaching. The questions also delve into the pros and cons of learning from both types of teachers, allowing respondents to consider factors such as accent, cultural knowledge, teaching approach, and language proficiency. The last questions focus on identifying the key characteristics that respondents consider essential for any English teacher, regardless of their native language. These questions aim to uncover the common criteria that students value in their language teachers, such as effective communication skills, knowledge of grammar and vocabulary, accessibility, cultural sensitivity and the ability to create a favorable learning environment. The detailed list of questions is shown in the appendix.

The questionnaire was presented in English and Japanese, thereby removing any confusion by students who might have problems understanding the questions. In addition, the study does not assess proficiency in English language; therefore, grammatical and textual mistakes are overlooked during the analysis, except in instances where the intended message is ambiguous, leading to the exclusion of the data. The answers written in Japanese were translated by a bilingual Japanese speaker and checked by a native Japanese speaker.

Procedure

Participants were asked to complete a Google Form during their classes. Participants who agreed to take part were formally introduced to the study through a consent form before being asked to fill out a questionnaire, which required approximately 15 to 20 minutes to finish. The questionnaire utilized in the study ensured anonymity, with no additional identifiable information gathered from the participants.

Following the collection of data, a framework was developed through the identification of emerging themes. The design of the framework was characterized by its extensive and non-hierarchical nature, primarily serving the descriptive purpose of organizing the data based on themes. The gathered data were then quantified based on the themes outlined in the coding framework, which are detailed in the first column on the left-hand side of Table 1.

Results

Table 1*Students' perceptions of their teacher's characteristics*

N=57	Strongly Disagree	Disagree	Neither Agree nor disagree	Agree	Strongly Agree	Average Mean Score
My NEST...	1	2	3	4	5	
has a good command of the English language and pronunciation.	0	0	2 (3.5%)	5 (8.8%)	50 (87.7%)	4.84
is effective in explaining complex grammar concepts.	0	2 (3.5%)	10 (17.6%)	18 (31.6%)	27 (47.3%)	4.22
is approachable and encourages student participation.	0	0	2 (3.5%)	8 (14.1%)	47 (82.4%)	4.78
brings cultural insights and real-life experiences from English-speaking countries.	0	0	3 (5.3%)	12 (21%)	42 (73.7%)	4.68
has a deeper understanding of English literature and idiomatic expressions.	0	0	9 (15.8%)	11 (19.3%)	37 (64.9%)	4.52
I feel comfortable asking questions to my teacher in class.	0	4 (7%)	9 (15.8%)	16 (28.1%)	28 (49.1%)	4.19
covers the points I think are important to learn English.	0	0	4 (7%)	18 (31.6%)	35 (61.4%)	4.54
N=167	Strongly Disagree	Disagree	Neither Agree nor disagree	Agree	Strongly Agree	Average Mean Score
My N-NEST...	1	2	3	4	5	

has a good command of the English language and pronunciation.	1 (0.6%)	0	1 (0.6%)	19 (11.4%)	146 (87.4%)	4.85
is effective in explaining complex grammar concepts.	0	5 (3%)	17 (10.2%)	59 (35.3%)	86 (51.5%)	4.35
is approachable and encourages student participation.	0	1 (0.6%)	5 (3%)	30 (18%)	131 (78.4%)	4.74
brings cultural insights and real-life experiences from English-speaking countries.	1 (0.6%)	2 (1.2%)	16 (9.6%)	39 (23.3%)	109 (65.3%)	4.51
has a deeper understanding of English literature and idiomatic expressions.	1 (0.6%)	1 (0.6%)	6 (3.6%)	44 (26.3%)	115 (68.9%)	4.62
I feel comfortable asking questions to my teacher in class.	0	2 (1.2%)	19 (11.4%)	49 (29.3%)	97 (58.1%)	4.44
covers the points I think are important to learn English.	0	3 (1.8%)	8 (4.8%)	54 (32.3%)	102 (61.1%)	4.52

Table 2

Students' descriptions of their teacher's advantages

	Total	% compared to N (224)
With a NEST		
Authentic Pronunciation	128	57.1%
Listening skills	10	4.5%
Knowledge of foreign cultures	7	2.6%
With a N-NEST		

Pronunciation (as a negative)	44	19.6%
Local Cultural Understanding	11	4.9%
Ability to speak the same language as students (Japanese)	56	25%

Table 3

Students' classifications of whether or not their teacher is a NEST or N-NEST when in fact taught by a N-NEST

	Total	% compared to n (167)
Answered NEST	145	86.8%
Answered N-NEST	13	7.8%
Answered "not sure"	9	5.4%

Table 4

Students' answers regarding what type of teacher they would prefer

	Total	% compared to n (224)
Preferred NEST	164	73.2%
Preferred N-NEST	10	4.5%
Answered "both"	15	6.7%
No Preference	9	4%
No answer	32	14.3%

Findings and Discussion

Natural Pronunciation and Fluency

In general, the subjects placed particular importance on NESTs as an authentic model for pronunciation, with 128 students mentioning it as being the main benefit of having what they perceived to be a NEST. One such student mentioned, "I was able to experience real English pronunciation and culture." On the other hand, this was considered a negative point

when learning with a N-NEST, with 44 students mentioning that pronunciation was “unauthentic” or “difficult to understand.” These results are similar to those in a study conducted by Pacek (2005) at a British university where 43 international students were surveyed. Among them, 24 students emphasized the importance of clear pronunciation in language instructors, while only seven students specifically advocated for native-speaker pronunciation. One hundred and twenty-eight students in one way or another mentioned pronunciation as a positive when studying specifically under NESTs, while 44 students mentioned pronunciation as a negative when studying with N-NESTs. Some students expressed frustration with their NESTs and reported that “it was hard to keep up in class.” This indicates that while learners assert their high valuation of pronunciation that is deemed “authentic,” this assertion may present a paradoxical challenge for them, as they may encounter difficulties in deciphering English when spoken at a natural pace.

Another benefit that students mentioned regarding learning from a NEST was that they could improve their listening skills, with 10 students answering this was a priority for them. Finally, seven students answered that learning about the NEST’s culture was important to them when learning English.

Cultural Insights

When commenting about the benefits of having either a NEST or a N-NEST, many students mentioned how useful it was to learn about the teacher’s culture first-hand. A student commented, “(I) can learn about native speaker’s country culture,” while another listed “Experiencing real culture” as a positive point. In this study, six students mentioned specifically how learning about cultures different than theirs was important for them. The literature shows that access to differing cultural perspectives can benefit students. Tashev (2021) mentions in his study the importance of having a pluri-cultural perspective for students, as this approach would better prepare students for real-world communication that increasingly involves interactions with speakers from various backgrounds.

The idea of intercultural education is brought up in another study where Japanese students in an Australian university showed that their exposure to other cultures helped in their willingness to share personal thoughts and feelings and enhanced their bond with their Australian peers. This study also concluded that these interactions helped enhance empathy and the ability to understand and share their emotions. International programs are perceived as better when students can navigate foreign cultures, build relationships, and communicate

freely (Kudo & Simkin, 2003). Students don't necessarily have to go abroad to benefit from intercultural exposure. These benefits were also shown in a study conducted by Hofmeyr (2022), which argued that students who had participated in intercultural training tended to exhibit more positive contact strategies and better understanding of others. Students who didn't participate in these programs reported more feelings of anxiety compared to the ones who did. These studies strongly suggest that Japanese students who have knowledge of different cultures tend to excel in their communication, personal, and emotional skills. Furthermore, the results support the hypothesis that the presence of a multicultural academic staff, characterized by a variety of cultural heritages, may contribute to a more profound and enriching educational experience for students.

Teachers' Academic Qualities

Among similar studies, results show that students perceive both NESTs and N-NESTs to have different abilities when teaching. NESTs tend to focus more on pronunciation and active classes, while N-NESTs are perceived to be better organized and grammar-oriented (Xu & Xiao, 2023). Both groups had different methodologies that benefited the academic performance of students.

However, an interesting finding of the study was that, regardless of an instructor's native speaker status, certain academic qualities were important for the participants in our study. Seventy-nine students mentioned this particular aspect in their open responses when asked "In your opinion, what are the key qualities that an effective English teacher should possess, regardless of their native language?." Some of their answers were: "(teachers should) provide classes that are consistent with each student's English skills" and "Thinking about lesson content so that students can enjoy participating in class." Students also found it important that teachers should focus on methodologies that brought equity to the classes: "Being able to teach classes that match the English level of the students" and "Having an attitude of not leaving students who do not understand."

Fifty-six students mentioned the importance of being able to use Japanese when asking the teacher questions and when listening to the teacher's explanation of difficult English grammar points, and 11 stated that having the same cultural background as their instructor is important. Clearly, having the ability to communicate in the students' L1 is seen as a beneficial characteristic of an instructor. This reflects Medgyes' (1992) conclusion that an ideal NEST is one who has achieved a high level of fluency in the students' L1.

Among other comments that focused on the teachers' academic abilities, students noted that teachers should, "Provide classes that are consistent with each student's English skills, but never let them speak a language other than English," "They should be willing to explain concepts and simplify them when needed," and "they should not just rely on a textbook to teach English, but also use other activities like videos and group work to teach." All these comments highlight how much students value their teacher's ability to communicate effectively in the classroom.

Understanding of Local Learning Challenges and L2

Traditionally, in Japanese education, teachers deliver their classes in the form of a lecture, which emphasizes a "one-way" style of teaching whereby the instructor's role is to deliver information and the student's role is to absorb it. Japanese students assume that when they speak, they are actually hindering their classmates' learning and disrupting the teacher's instruction (Sasaki & Ortlieb, 2017). Compared to other styles of teaching, in which students are more involved and tend to be more active, Japanese students used to this "one-way" method often have problems adapting to a more interactive classroom dynamic. This sometimes leads to confusion for beginner teachers who arrive in Japan and expect students to be responsive and active. Maher & King (2022) commented in a study that students who are silent tend to have a negative self-evaluation and self-doubt about their language proficiency.

Perhaps unsurprisingly, when participants in our study expressed their opinions on this subject, some expressed that, "(teachers should) empathize with us." Another student commented, "(I) can ask questions easily." On the other hand, a particular student said: "It is difficult for me to be able to express my opinions to my teacher." These comments may reflect the different teaching styles between non-Japanese teachers and Japanese teachers. Successful teachers should be able to recognize these differences and address them adequately in order to make students feel more comfortable and confident, which will allow them to freely speak their minds. These findings suggest that it is important for teachers to be familiar with the culture and language of the country they are teaching regardless of their status as a NEST or N-NEST.

Students' Focus on the Teachers' Personal Qualities

When answering the same open-ended question, “In your opinion, what are the key qualities that an effective English teacher should possess, regardless of their native language?”, 63 students mentioned some sort of personality skill as important. Out of these answers, 44 students specifically mentioned “Friendly” as their most important factor when studying under a foreign teacher. Some of the other personality qualities that students felt a teacher should have were “considerate,” “good attitude,” “interesting,” and “culturally aware.”

A study by Matsumoto (2017) shows that 36.8% of university students considered their teacher’s personality as a strong factor influencing their motivation and performance in class. While this result did not represent a significant influence when the teacher was a NEST, it was an especially significant factor when the teacher was a N-NEST. In other words, whether or not a N-NEST possesses a natural, approachable attitude and egalitarian approach to education in the classroom can have a significant impact on student engagement in learning.

Students’ Perception of the Instructor Being a NEST or N-NEST

One complication that showed up in our research was the fact that many students couldn’t tell if their instructor was either a NEST or a N-NEST. According to the data shown in Table 3, out of a total of 167 students who had a N-NEST, where 58 were female and 109 were male, 47 female participants and 98 male participants answered incorrectly when asked if their instructor was a Native or a Non-Native English Speaker. Only five female students (8.6%) and eight male students (7.3%) answered correctly.

The data suggests that students may not be aware of the pronunciation and cultural differences between a NEST and a N-NEST, which supports the idea that students' preference might be more influenced by the instructor's other non-tangibles – such as personality, preparedness, rapport, and class activities – rather than by their instructor’s nationality.

Students’ Preference for Either NEST or N-NEST

Regardless of whether the instructor is a NEST or a N-NEST, students are able to recognize strengths that both groups can bring to the classroom. Among studies done in other countries with similar groups, the results pointed to a similar conclusion, where students understood the value of having teachers from various backgrounds (Arnoud et al., 2023; Deng et al., 2023; Phothongsunan, 2017; Xu & Xiao, 2023). Both the subjects and the results

of these studies highlight that methodology, pedagogical knowledge, attitude, and preparedness are more valued in teachers than their countries of origin, since both NESTs and N-NESTs have significant benefits. Similarly, in our study, the overall response from the students seemed to be focused mostly on the instructor's pronunciation, personality, way of teaching, methodologies, class activities, and rapport with the students. According to Table 4, out of 168 students who answered that they would prefer having an instructor being a NEST, 62 were female and 102 were male, while 10 students preferred having a N-NEST with an even split between female students and male students. In the case of "both," only nine male students chose this option. Finally, nine students, of whom three were female and six were male, had no particular preference. Thirty-two students did not answer this question.

Similar findings were obtained with studies done in other countries (Arnoud et al., 2023), where students found no significant difference in their instructor's effectiveness regardless of nationality, meaning that whether a teacher was native or non-native did not have an impact on how students rated their performances. Instead, students put more emphasis on the instructor's pedagogical prowess than on their nationality. Another study done in seven Chinese universities also concluded that students' perceptions of NESTs and N-NESTs are shaped by the teaching methodologies utilized and the instructors' ability to tailor their approaches to teaching English (Xu & Xiao, 2023).

Ahmed and Osam (2022) noted in their study that the subjects of their study had differing views on what constituted effective teaching qualities. Whereas NESTs were preferred for their personal attributes such as caring, patience, and engagement, those who preferred N-NESTs did so for attributes related to professional competencies including knowledge, teaching skills, and professionalism.

Limitations

The primary limitation of this study is the fact that answers were obtained only from students at one university. Due to the nature of the majors taught at the university, other majors that might have more international exposure are lacking in this study. Another limitation is the sample size. Having a larger data set would undoubtedly yield more precise answers to the questions asked. Specifically, the group of students who answered the questionnaires given by a N-NEST is a larger sample than the group who answered the same questionnaire given by a NEST. In response to this particular problem, we calculate the results as percentages related to the total number of students who answered the questionnaire.

An additional constraint is that the sample is specifically selected from a tertiary institution, thus limiting its generalizability to different language learning environments. The study's reliance on a single instrument for data collection is noted. While this is a prevalent approach in research on this subject, the inclusion of analysis with alternative data collection methods could have enhanced the richness and detail of the information. Finally, the students from whom the data were collected had two different instructors: a main English teacher who has class with them four days a week and a Japanese English teacher whom they see once a week. It is possible that students identified their Japanese English teacher as a N-NEST and their main English instructor as a NEST regardless of their native speaker status. Survey results were omitted in cases where there was an obvious misunderstanding. Further research that uses a group of students who do not have Japanese teachers of English at the time may yield less ambiguous results.

Conclusion

This study attempted to survey the perceptions of students toward NESTs and N-NESTs in a Japanese university context using a mixed-method design. The Likert-scale survey findings suggest that students may find N-NEST teachers more approachable than their NEST counterparts, while the data obtained from students' short answer responses to the survey showed the perceived advantages and disadvantages of being taught by either a NEST or a N-NEST. Students overwhelmingly chose having a NEST over a N-NEST, while at the same time failing to identify which group their own teachers fell into. Limitations of the study included a somewhat small sample size as well as homogeneity. Also, it should be noted that although efforts were made to make it clear that this was a study focused on their English instructors and not on their Japanese English teachers, there is a possibility that students might have ignored this instruction and confused their Japanese teachers of English as N-NEST, giving their N-NEST a grade and feedback meant for their Japanese teacher. There is no conclusive evidence to prove that students did or did not confuse their Japanese English instructor with their N-NEST. Further studies are needed to conclusively determine this possibility.

The result of this and other similar studies suggest that stakeholders in education, including recruiters and policymakers, should consider the diverse strengths of both NESTs and N-NESTs when hiring teachers so that institutions and students can benefit from the diverse strengths that both NESTs and N-NESTs bring. Further study would include

responses from other universities in Japan. Also, a future qualitative study on Japanese university students' reasons for finding N-NESTs to be more approachable could offer information regarding this trend.

References

- Ahmed, S. K., & Osam, N. (2022). Students' and instructors' attitudes and perceptions towards Native and Non-native EFL teachers in the Kurdistan region of Iraq. *Arab World English Journal*, 13(4), 115–139. <https://doi.org/10.31235/osf.io/xn5uc>
- Benke, E., & Medgyes, P. (2005). Differences in teaching behaviour between native and non-native speaker teachers: As seen by the learners. In E. Llurda (Ed.), *Non-native language teachers: Perceptions, challenges and contributions to the profession* (pp. 195–215). Springer New York.
- Cazden, C. B. (2001). *Classroom discourse: the language of teaching and learning* (2nd ed.). Heinemann.
- Chiba, R., Matsuura, H., & Yamamoto, A. (1995). Japanese attitudes toward English accents. *World Englishes*, 14(1), 77–86. <https://doi.org/10.1111/j.1467-971X.1995.tb00341.x>
- Cumming, B., & Matsumoto, M. (2018). Japanese high school and university students learning English: Differences in their perceptions of teachers as motivators. *The Asian Journal of Applied Linguistics*, 5(2), 217–229.
- Deng, L., Zhang, L. J., & Mohamed, N. (2023). Exploring native and non-native English speaker teachers' perceptions of English teacher qualities and their students' responses. *Frontiers in Psychology*, 14, 1–9. <https://doi.org/10.3389/fpsyg.2023.1175379>
- Egitim, S., & Garcia, T. (2021). Japanese university students' perceptions of foreign English teachers. *English Language Teaching*, 14(5), 13–22. <https://doi.org/10.5539/elt.v14n5p13>
- Esaki, Y., & Shintani, R. (2010). English education in Japan's elementary schools: Reforms and challenges. *Asia Pacific World*, 1(1), 133–154. <https://doi.org/10.3167/APW.2010010109>
- Ghane, M. H., & Razmi, M. H. (2023). Exploring the effectiveness of native and non-native English teachers on EFL learners' accuracy, fluency, and complexity in speaking. *Education Research International*, 2023(1), 1–8. <https://doi.org/10.1155/2023/4011255>
- Hakimian, L. (2024). TEFL students' perceptions of native and non-native EFL school teachers and university lecturers: A case study. *European Journal of Teaching and Education*, 6(2), 12–30. <https://doi.org/10.33422/ejte.v6i2.1224>
- Hofmeyr, A. S. (2022). Japanese university student responses to intercultural scenarios: Contact strategies and perceptions of otherness. *International Journal of Intercultural Relations*, 90, 73–85. <https://doi.org/10.1016/j.ijintrel.2022.07.008>
- Johannes, A. A. (2012). Team teaching in Japan from the perspectives of the ALTs, the JTEs, and the students. *TEFLIN Journal*, 23(2), 165–182. <https://doi.org/10.15639/teflinjournal.v23i2/165-182>

- Kasai, M., Lee, J. A., & Kim, S. (2011). Secondary EFL students' perceptions of native and nonnative English-speaking teachers in Japan and Korea. *Asian EFL Journal*, 13(3), 272–300.
- Kudo, K., & Simkin, K. A. (2003). Intercultural friendship formation: The case of Japanese students at an Australian university. *Journal of Intercultural Studies*, 24(2), 91–114. <https://doi.org/10.1080/0725686032000165351>
- Lai, H. Y. T. (2023). The perceptions of university English teachers in Japan about teaching English as an international language. *World Journal of English Language*, 13(7), 100–107. <https://doi.org/10.5430/wjel.v13n7p100>
- Lasagabaster, D., & Sierra, J. M. (2002). University students' perceptions of native and non-native speaker teachers of English. *Language Awareness*, 11(2), 132–142. <https://doi.org/10.1080/09658410208667051>
- Ling, C. Y., & Braine, G. (2007). The attitudes of university students towards non-native speakers English teachers in Hong Kong. *REL C Journal*, 38(3), 257–277. <https://doi.org/10.1177/0033688207085847>
- Mahboob, A. (2004). Native or non-native: What do the students think? In L. D. Kamhi-Stein (Ed.), *Learning and teaching from experience* (pp. 121-147). University of Michigan Press.
- Mahboob, A. (2005). Beyond the native speaker in TESOL. *Culture, Context, & Communication*, 30, 60–93.
- Maher, K., & King, J. (2022). 'The silence kills me.': 'Silence' as a trigger of speaking-related anxiety in the English-medium classroom. *English Teaching & Learning*, 46(3), 213–234. <https://doi.org/10.1007/s42321-022-00119-4>
- Matsumoto, M. (2017). Japanese EFL students' perceptions of native and non-native teachers of English as a motivational factor. *Journal of Modern Languages*, 27, 1–23. Retrieved September 15, 2024, from <http://jummec.um.edu.my/index.php/JML/article/view/6845>
- Medgyes, P. (1992). Native or non-native: Who's worth more? *ELT Journal*, 46(4), 340–349. <https://doi.org/10.1093/elt/46.4.340>
- Moussu, L. M. (2006). *Native and nonnative English-speaking English as a second language teachers: student attitudes, teacher self-perceptions, and intensive English administrator beliefs and practices* [Doctoral thesis, Purdue University]. Education Resources Information Center. <https://eric.ed.gov/?id=ED492599>
- Murtada, A. Y., & Alsuhaibani, Y. A. (2023). The validity of hiring native and non-native English-speaking instructors: Professional observers', learners', and programme administrators' perceptions. *Journal of Language Teaching and Research*, 14(3), 689–700. <https://doi.org/10.17507/jltr.1403.16>

- Nagatomo, D. H. (2019). *Identity, gender and teaching English in Japan*. Multilingual Matters.
- Okada, T. (2023). Outsider teachers? Filipino teachers' reflections on English teaching and raising intercultural awareness in Japan. In G. P. Glasgow (Ed.), *Multiculturalism, language, and race in English education in Japan: Agency, pedagogy, and reckoning* (pp. 204-225). Candlin & Mynard ePublishing. <https://doi.org/10.47908/26/8>
- Oude Groote Beverborg, A., & Müller, B. C. N. (2024). Students' evaluations of native and non-native teachers in higher education. *Assessment & Evaluation in Higher Education*, 49(3), 275–287. <https://doi.org/10.1080/02602938.2023.2213422>
- Pacek, D. (2005). 'Personality not nationality': Foreign students' perceptions of a non-native speaker lecturer of English at a British university. In E. Llurda (Ed.), *Non-native language teachers: Perceptions, challenges and contributions to the profession* (pp. 243-262). Springer.
- Phothongsunan, S. (2017). Delving into international students' attitudes to NESTs and NNESTs. *3L, Language, Linguistics, Literature*, 23(4), 15–28. <http://doi.org/10.17576/3L-2017-2304-02>
- Rasyid, F., Ibna Seraj, P. M., Ghofur, A., & Asrifan, A. (2023). Students' perception toward teaching strategies of native and nonnative English-speaking teachers: A case study in Indonesia. *Education Research International*, 2023(1), 1–9. <https://doi.org/10.1155/2023/7827917>
- Sasaki, Y., & Ortlieb, E. (2017). Investigating why Japanese students remain silent in Australian university classrooms: The influences of culture and identity. *Journal of Asian Pacific Communication*, 27(1), 85–98. <https://doi.org/10.1075/japc.27.1.05sas>
- Selvi, A. F. (2014). Myths and misconceptions about nonnative English speakers in the TESOL (NNEST) movement. *TESOL Journal*, 5(3), 573–611. <https://doi.org/10.1002/tesj.158>
- Tajino, A., & Tajino, Y. (2000). Native and non-native: What can they offer? Lessons from team-teaching in Japan. *ELT Journal*, 54(1), 3–11. <https://doi.org/10.1093/elt/54.1.3>
- Tassev, V. V. (2021). Undergraduate university students' views regarding intelligibility and comprehensibility of the pronunciation of native English teachers' (NETs) versus non-native English teachers' (NNESTs') pronunciation - A case-study at Assumption University of Thailand (AU). *The New English Teacher*, 15(1), 24–35.
- Todd, R. W., & Pojanapunya, P. (2009). Implicit attitudes towards native and non-native speaker teachers. *System*, 37(1), 23–33. <https://doi.org/10.1016/j.system.2008.08.002>
- Toya, M. (2023). Role of study abroad in the path to teaching English in English among Japanese teachers. *SEACE Official Conference Proceedings. The International Academic Forum (IAFOR). Singapore*. 2435–5240. <https://doi.org/10.22492/issn.2435-5240.2023.48>

- Tsou, S. Y., & Chen, Y. L. (2017). EFL college students' perceptions toward native and non-native English speaking teachers. *American Journal of Educational Research*, 5(12), 1182–1190. <https://doi.org/10.12691/education-5-12-2>
- Walker, E. (2001). Roles of native-speaker English teachers (NETs) in Hong Kong Secondary Schools. *Asia Pacific Journal of Language in Education*, 4(2), 51–77.
- Walkinshaw, I., & Duong, O. T. H. (2012). Native-and non-native speaking English teachers in Vietnam: Weighing the benefits. *TESL-EJ*, 16(3), 1–17.
- Walkinshaw, I., & Oanh, D. H. (2014). Native and non-native English language teachers: Student perceptions in Vietnam and Japan. *Sage Open*, 4(2), 1–9. <https://doi.org/10.1177/2158244014534451>
- Walsh, S. (2002). Construction or obstruction: Teacher talk and learner involvement in the EFL classroom. *Language Teaching Research*, 6(1), 3–23. <https://doi.org/10.1191/1362168802lr095oa>
- Wang, L. Y. (2012). Moving towards the transition: Non-native EFL teachers' perception of native-speaker norms and responses to varieties of English in the era of global spread of English. *Asian EFL Journal*, 14(2), 46–78.
- Xu L. & Xiao Y. (2023). Students' perceptions of native/non-native English-speaking EMI teachers: Are NS teachers better than NNS teachers? *Frontiers in Education*, 7, 1–10. <https://doi.org/10.3389/educ.2022.1082600>
- Yazawa, O. (2017). Students' perception of native English-speaking teachers and Japanese teachers of English: The effect on students' self-efficacy and emotional state. *Eruditi*, 1(3), 61–72.

Appendix

Questionnaire

The following questions are the introductory questions:

- How long have you been learning English?
- Have you learned English with a Native teacher before? (Teacher from America, England, Australia, New Zealand, Ireland, Scotland or Canada)
- Have you learned English with a non-Native teacher before? (Any other country except America, England, Australia, New Zealand, Ireland, Scotland or Canada)
- Is your current English teacher Native or non-Native?
- What do you think is important to learn while learning English?
- Do you have a preference between learning English from a Native speaker, a non-Native Speaker?

The following questions are to be answered on a scale from 1 to 5:

- My current teacher covers the points I think are important to learn English.
- My current teacher is effective in explaining complex grammar concepts.
- My current teacher has a good command of the English language and pronunciation.
- My current teacher brings cultural insights and real-life experiences from English-speaking countries.
- My current teacher has a deeper understanding of English literature and idiomatic expressions.
- I feel comfortable asking questions to my teacher in class.

The following are open-ended questions:

- What do you think are advantages and disadvantages of learning from a Native English speaker?
- What do you think are advantages and disadvantages of learning from a Foreign non-Native English speaker?
- In your opinion, what are the key qualities that an effective English teacher should possess, regardless of their native language?

A Readability Based Comparative Study of Four Corners 1 & TOEIC Skills 1

Joshua Trued, Asia University

Abstract

This study aims to evaluate the vocabulary and reading sections of the textbooks *Four Corners 1 (FCI)* and *TOEIC Skills 1 (TSI)* concerning their readability levels and vocabulary profile. Utilizing tools such as the Content Vocabulary Level Analyzer and Coh-Metrix, the analysis focused on aspects such as the texts' word difficulty, word concreteness, and syntactic simplicity. The findings indicate that while *FCI* is appropriate for low-level students, *TSI* presents a mismatch in both vocabulary and reading difficulty. If used together, *TSI* may overwhelm lower-level students. These results suggest a need for careful consideration of materials to ensure effective learning outcomes.

Introduction

Ideally, an English as a Foreign Language (EFL) textbook should align with curriculum objectives, facilitate teaching and learning, and cover all language skills (Hanifa, 2018). Also, it should provide engaging content, clear instructions, and culturally appropriate topics (Hanifa, 2018). However, textbooks may have such shortcomings as a lack of pair/group work exercises or an insufficient amount of vocabulary reinforcement. The selection of EFL textbooks is crucial, with factors including curriculum alignment, content quality, and price influencing decisions (Wuttisrisiriporn et al., 2020). Some objective evaluation of textbooks is recommended before use to ensure their suitability for the EFL classroom and to maximize teaching and learning outcomes (Wuttisrisiriporn et al., 2020).

One major factor in assessing a textbook's effectiveness is readability (Gunning, 2003), which is defined as the ease and speed with which a reader can comprehend a text. It encompasses various characteristics that contribute to comprehension, including legibility, syntactic and semantic difficulty, and text organization (Kane et al., 2006). Researchers have identified numerous variables influencing readability, with one study finding as many as 228 factors categorized into content, style, format, and organizational features (Dreyer, 1984; Kane et al., 2006). At any rate, readability formulae provide instructors with a way to objectively compare potential textbooks.

Vocabulary taught in an academic course usually derives from the textbook used, and even in instances where additional resources are incorporated, it is reasonable to assume that the complexity of these supplementary materials corresponds with that of the primary textbook. Although publishers often provide guidelines regarding the difficulty level associated with their textbooks, empirical outcomes can exhibit variability, suggesting that educators should possess a mechanism for an independent evaluation of this criterion.

This researcher's Freshman English classes require the use of *Four Corners 1 (FCI)* and *TOEIC Skills (TSI)* together. This paper aims to answer the following research question: 1. How do *FCI* and *TSI*, whose respective publishers state a slightly different target learner level, actually compare in terms of their readability scores and vocabulary profile?

Literature Review

Text readability, which encompasses all characteristics contributing to ease of comprehension, is influenced by numerous factors including content, style, format, and organization (Kane et al., 2006). It is crucial for ensuring that texts, particularly assessment

tasks, are at a level appropriate to their intended readers (Wray & Dahlia, 2013). The number of available readability formulae is high, but in general, such formulae focus on two key areas of a text: the difficulty of words and sentences (DuBay, 2004). Sentence difficulty is measured in terms of how many words a sentence contains, while word difficulty is measured by how long the word is or how many syllables it has.

The Flesch-Kincaid Reading Ease and Flesch-Kincaid Grade Level (both developed in 1948) are standard metrics for assessing readability. The Flesch-Kincaid Grade Level indicates the requisite United States education level for text comprehension. For instance, a text rated at Grade Level 5 necessitates approximately five years of educational experience for comprehension.

This metric was originally created for L1 users (i.e., students whose first language is the language being learned, in this case, English). Because word and sentence length, which most readability formulae use to measure difficulty, do not necessarily indicate comprehensibility, it can be said that the knowledge gap regarding vocabulary between an L1 user and an L2 user (i.e., a student whose first language is not the language being learned) may be substantial considering, for example, that Japanese students would likely have less knowledge of Latin-based words than a Spanish- or Italian-speaking student (Abdollahi-Guilani, 2022). However, this readability formula is widely used in the field and is even available as a built-in feature in Microsoft Word. Regarding it being used in lieu of other formulae, there is little difference among them concerning their validity (Zamanian & Heydari, 2012).

Several readability formulae have been created specifically for L2 users. These include the CEFR-based Vocabulary Analyzer (CVLA) and Coh-Metrix. The CVLA is an online tool created by researcher Satoru Uchida that analyzes text and determines its level according to the Japanese version of the Common European Framework of Reference (CEFR-J). The CEFR-J was developed to better assess English proficiency among Japanese learners (Tono, 2019). It expanded the original CEFR by introducing sub-levels within A1, A2, B1, and B2, aiming to discriminate finer proficiency differences (Runnels, 2013). The CEFR-J project, launched in 2008, created descriptors for 10 sub-levels and developed related resources, including wordlists and e-learning tools (Tono, 2019). While the CEFR has six levels, with A1 being the lowest and C2 being the most challenging, the CEFR-J includes 12 levels, with PreA being the lowest and C2 the highest (see Figure 2) (Uchida & Negishi, 2018).

Another readability formula created specifically for L2 users is the Coh-Metrix tool. This is a web-based tool designed to automatically evaluate text cohesion and coherence across multiple linguistic levels (McNamara et al., 2010). It analyzes word characteristics, sentence features, and discourse relationships, providing a comprehensive set of indices (McNamara & Graesser, 2012). Coh-Metrix has been validated as a measure of cohesion, successfully distinguishing between high and low-cohesion texts in published psychology studies (McNamara et al., 2010). It offers advantages over traditional readability measures by going further than mere sentence and word length and assesses text at various language and discourse levels (McNamara & Graesser, 2012).

Methodology

Textbooks

This study assessed *FCI* and *TSI*. *FCI* is appropriate for students with a TOEIC score of more than 120 and a CEFR level of A1-A2. Somewhat contrastingly, *TSI* is appropriate for students with a TOEIC score of 350-400. The corresponding CEFR level for this is A2 (see Figure 1 and Figure 2).

Figure 1

Four Corners Level 1 in relation to the CEFR

Four Corners Second Edition and the Common European Framework of Reference

The CEFR was used as a point of reference throughout the development of the *Four Corners Second Edition* series and informed the choice of language content and activity. The table below shows how *Four Corners Second Edition* correlates with the Council of Europe's levels and with some major international examinations.

Four Corners Second Edition	CEFR	Council of Europe	Cambridge ESOL	IELTS	TOEFL iBT	TOEIC
Level 1	A1/A2	Breakthrough				120+

Source: https://www.cambridge.org/jp/files/8515/3924/6166/CEFR_TE_Level_1.pdf

Figure 2

Breakdown of CEFR-J levels

CEFR		CEFRとCEFR-Jの関係	CEFR-J
Proficient User 【熟達した使用者】 <small>様々な状況で英語を意のままに用いることができ、場面に合った言葉づかいや文体を選択できる。</small>	C2		C2
	C1		C1
Independent User 【自立した使用者】 <small>日常生活や学習、仕事などで必要な英語を理解し、ある程度流暢に自分を表現することができる。</small>	B2	投野由紀夫教授 (東京外国語大学) を中心に CEFR をもとに、 日本の英語教育の 実状に合うよう細分化	B2.2
			B2.1
	B1		B1.2
			B1.1
Basic User 【基礎段階の使用者】 <small>身近なことやなじみのあるトピックであれば、日常生活で使われる英語を理解し、用いることができる。</small>	A2		A2.2
			A2.1
			A1.3
	A1		A1.2
			A1.1
			PreA1

※レベルを示す表現は、Z会グループ「基礎学力総合研究所」の定義によるものです。

Source: <https://logos.edu.iwate-u.ac.jp/jhoffice/wp-content/uploads/sites/6/2019/04/CEFR-J.png>

Procedure

Material used for this study included ten units of *FCI* and ten units of *TSI*.

Vocabulary and reading passages from both books were scanned and input into a Word file so that they could be easily imported into the CVLA and the Coh-Metrix tools. In *FCI*, vocabulary is introduced in two out of four sections (A and C). *TSI* includes a section called “Focus on Vocabulary” in part A and part B of every unit. These were the only areas from which vocabulary terms were scanned and analyzed.

To gain a better insight as to the appropriateness of *FCI* and *TSI* vocabulary in a Japanese context, Uchida’s CVLA was utilized (<https://cvla.langedu.jp/>). CVLA is a program which allows users to assess text metrics such as the approximate CEFR-J level, verbs per sentence, average word difficulty, and the ratio of CEFR B words to CEFR A words. The tool also classifies each word according to its CEFR level. For this study, each unit’s vocabulary from both textbooks were assessed according to the corresponding CEFR level. Furthermore, the number of words for each classification is shown. The CVLA was also used to assess the

reading portions of each text. The CEFR-J level, verbs per sentence, and average word difficulty were recorded.

Regarding reading, *FCI* typically has reading material in section D of each unit, and *TSI* has a section called “Reading: Text Completion” in each unit. (In principle, reading material from the two textbooks was entered into a word processing application in a manner that accurately reflected the presentation of the text as it appeared in the source material.)

The Coh-Metrix tool (<https://soletlab.adaptiveliteracy.com:8443/Coh-MetrixResearch.aspx>) was used to measure cohesion and text difficulty of *FCI* and *TSI* reading material by analyzing factors such as narrativity, syntactic simplicity (SS), word concreteness (WC), referential cohesion, and deep cohesion. The tool also provides the corresponding Flesch-Kincaid grade level for each text. Because of the relatively short length of the reading sections, this study only focused on the SS, WC, and the Flesch-Kincaid Grade Level. SS is measured utilizing various metrics, including the mean number of clauses per sentence, the total word count per sentence, and the word count prior to the principal verb of the main clause. WC measures the abstractness of words in a text. If words can be experienced using one or more of the five senses, they would not be considered to be abstract. The higher the score for WC, the easier it is for a text to be comprehended (Graesser et al., 2014). For both SS and WC, a higher percentage indicates a more comprehensible text.

Results

Vocabulary Analysis

The vocabulary across units in both *FCI* and *TSI* were analyzed using the CVLA tool to determine the CEFR level of each word (see Table 1 and Table 2).

Table 1
Vocabulary according to CEFR level from each unit of FCI.

Unit 1	A1: 14	A2: 4	B1: 2		
Unit 2	A1: 15				N/A: 7 (various nationalities)
Unit 3	A1: 19	A2: 8	B1: 4		N/A: 2 (eraser, blouse)
Unit 4	A1: 22	A2: 5	B1: 1		
Unit 5	A1: 25	A2: 2	B1: 3	B2: 1	N/A: 1 (apps)
Unit 6	A1: 17	A2: 5	B1: 2	B2: 1	N/A: 2 (French, attendant)
Unit 7	A1: 18	A2: 4	B1: 1		N/A: 2 (noodles, dumpling, sushi, taco)

Unit 8	A1: 21	A2: 4			N/A: 2 (aquarium, newsstand)
Unit 9	A1: 15	A2: 12			N/A: 2 (Italian, tutor)
Unit 10	A1: 31	B1: 2	B2: 1		N/A: 1 (grocery)
Unit 11	A1: 18	A2: 6	B1: 1		
Unit 12	A1: 10	A2: 4	B1: 2	B2: 1	

Table 2

Vocabulary according to CEFR level from each unit of TS1.

Unit 1	A1:5	A2:17	B1:10	B2:5	C1:1	N/A: 1 (janitor)
Unit 2	A1:11	A2:6				
Unit 3	A1:1	A2:9	B1:8	B2:2		
Unit 4	A1:2	A2:5	B1:1	B2:7		N/A: 3 (ready, license, cancelation)
Unit 5	A1:10	A2:11	B1:5	B2:4		N/A: 2 (repairman, warehouse)
Unit 6	A1:8	A2:4	B1:5	B2:1		N/A: 2 (data, log)
Unit 7	A1:7	A2:4	B1:2	B2:1		N/A:2 (shipment, invoice)
Unit 8	A1:7	A2:14	B1:9			N/A:1 (allergy)
Unit 9	A1:4	A2:10	B1:3	B2: 3		N/A:3 (closet, plenty, wireless)
Unit 10	A1:2	A2:3	B1:6	B2: 3		N/A:2 (accounting, advisor)

FCI

The majority of the vocabulary in *FCI* falls within the PreA1-A1 level, indicating that the textbook is well-suited for beginners. For example, Unit 1 contains 14 A1 words, four A2 words, and two B1 words, with an AWD of 1.4, corresponding to the A2.1 level. Across units, the CEFR-J levels generally ranged between PreA1 and A2.2, with minimal variation, showing consistency in difficulty. This indicates that *FCI* maintains a steady progression suitable for beginner learners.

TS1

Vocabulary in *TS1* was more variable than that in *FCI*, with several units containing higher-level words. Unit 1, for example, has five A1 words, 17 A2 words, and five B2 words. The AWD in *TS1* is consistently higher than in *FCI*, with Unit 1 having an AWD of 2.41 (C2 level). Other units (Unit 4 in particular) showed an AWD of 2.56 (C2 level), reflecting a significantly higher difficulty.

Reading Section Analysis

The reading sections of both textbooks were analyzed using the CVLA and Coh-Metrix tools (see Table 3 and Table 4).

Table 3

Reading section data from each unit of FC1.

	Verbs per Sentence	Average Word Difficulty	Flesch-Kincaid Grade:	SS:	WC:	CEFR-J Level:
Unit 1 Reading	1.05 (PreA1)	1.23 (A1.1)	7th grade (Fairly easy to read)	89%	12%	PreA1
Unit 2 Reading	1.11 (PreA1)	1.23 (A1.1)	6th grade (Easy to read)	56%	19%	PreA1
Unit 3 Reading	0.94 (PreA1)	1.48 (A2.2)	7th grade (Fairly easy to read)	56%	45%	A1.2
Unit 4 Reading	1.05 (PreA1)	1.08 (PreA1)	6th grade (Easy to read)	86%	27%	PreA1
Unit 5 Reading	1.45 (A1.2)	1.24 (A1.1)	6th grade (Easy to read)	92%	100%	A1.1
Unit 6 Reading	1.83 (A2.1)	1.33 (A1.3)	8th/9th grade (Plain English)	46%	100%	A1.3
Unit 7 Reading	1.5 (A1.2)	1.24 (A1.1)	6th grade (Easy to read)	89%	97%	A1.1
Unit 8 Reading	1.82 (A2.1)	1.32 (A1.3)	6th grade (Easy to read)	52%	84%	A1.2
Unit 9 Reading	1.5 (A1.2)	1.18 (PreA1)	6th grade (Easy to read)	73%	49%	PreA1
Unit 10 Reading	1.55 (A1.3)	1.13 (PreA1)	5th grade (Very easy to read)	73%	34%	PreA1
Unit 11 Reading	1.32 (A1.1)	1.26 (A1.1)	6th grade (Easy to read)	70%	56%	PreA1
Unit 12 Reading	1.88 (A2.1)	1.18 (PreA1)	7th grade (Fairly easy to read)	49%	99%	A1.2

Table 4

Reading section data from each unit of TS1.

	Verbs per Sentence	Average Word Difficulty	Flesch-Kincaid Grade:	SS:	WC:	CEFR-J Level:
Unit 1A Reading	2 (A2.2)	1.48 (A2.2)	7th grade (Fairly easy to read)	40%	87%	A2.2
Unit 1B Reading	4.67 (C2)	1.7 (B2.1)	12th grade (Fairly difficult to read)	5%	98%	C1
Unit 2A Reading	3.75 C2	1.71 B2.2	7th grade (Fairly easy to read)	1%	18%	B2.2
Unit 2B Reading	1.8 A2.1	1.48 A2.2	6th grade (Easy to read)	58%	10%	A1.3
Unit 3A Reading	2.57 B1.2	1.47 A2.2	6th grade (Easy to read)	63%	51%	A2.2

Unit 3B Reading	1.53 A1.3	1.37 A2.1	7th grade (Fairly easy to read)	92%	8%	A1.2
Unit 4A Reading	3.13 C1	1.43 A2.1	6th grade (Easy to read)	37%	14%	A2.2
Unit 4B Reading	2.22 B1.1	1.7 B2.1	8th & 9th Grade (Plain English)	91%	41%	B1.1
Unit 5A Reading	2.82 B2.1	1.42 A2.1	7th grade (Fairly easy to read)	61%	27%	A2.2
Unit 5B Reading	2 A2.2	1.28 A1.2	6th grade (Easy to read)	77%	27%	A1.2
Unit 6A Reading	2.56 B1.2	1.7 B2.1	7th grade (Fairly easy to read)	96%	39%	B1.2
Unit 6B Reading	2 A2.2	1.5 B1.1	7th grade (Fairly easy to read)	96%	95%	A2.1
Unit 7A Reading	1.88 A2.1	1.66 B2.1	6th grade (Easy to read)	73%	78%	A2.2
Unit 7B Reading	2.13 A2.2	1.45 A2.2	7th grade (Fairly easy to read)	90%	24%	A2.1
Unit 8A Reading	2.21 B1.1	1.51 B1.1	6th grade (Easy to read)	64%	85%	A2.2
Unit 8B Reading	2.8 B2.1	1.33 A1.3	6th grade (Easy to read)	54%	78%	A2.1
Unit 9A Reading	2.23 B1.1	1.61 B1.2	7th grade (Fairly easy to read)	53%	81%	A2.2
Unit 9B Reading	1 PreA1	1.51 B1.1	7th grade (Fairly easy to read)	66%	98%	A1.3
Unit 10A Reading	2.78 B2.1	1.75 B2.2	7th grade (Fairly easy to read)	86%	53%	B1.2
Unit 10B Reading	1.82 A2.1	1.83 C1	8th & 9th grade (Plain English)	82%	45%	B1.2

FCI

The reading passages in *FCI* displayed a strong alignment with the beginner level. For instance, the reading passage in Unit 1 scored a CEFR-J level of PreA1, with an average word difficulty of 1.23 (A1.1), and verbs per sentence were calculated at 1.05. Throughout the textbook, the reading passages ranged from PreA1 to A1.2, with most passages scoring low on the Flesch-Kincaid scale (6th to 7th grade). This consistency ensures that the reading material supports the intended learning outcomes for A1 learners.

FCI reading passages exhibit a high degree of WC throughout. This is reflected in the focus on everyday objects and beginner-level situations. For example, Unit 1 includes terms such as “actor,” “musician,” and “singer,” which are tangible and easily relatable for students at the PreA1-A2 level. Similarly, Unit 7, which deals with foods like “apples,” “bananas,”

and “dumplings,” maintains a strong focus on concrete nouns and simple actions. This emphasis on concrete vocabulary aligns with the textbook's overall goal of providing accessible content to beginner learners who may struggle with abstract language. This progression remains relatively consistent across the units, allowing students to gradually build their understanding of more complex vocabulary without being overwhelmed by abstract concepts.

The progression of syntactic complexity is minimal, with most units maintaining relatively simple sentence structures. For instance, Unit 1 has a verbs-per-sentence ratio of 1.05, which aligns with its PreA1 level. Even in later units, such as Unit 7 and Unit 8, the verbs-per-sentence ratio remains low (1.5 and 1.82, respectively), indicating that the textbook continues to prioritize syntactic simplicity to aid comprehension.

TSI

In contrast to *FCI*, the reading passages in *TSI* presented greater difficulty and variability. For example, Unit 1B's reading passage scored a CEFR-J level of C1, with an average word difficulty of 1.7 (B2.1) and a Flesch-Kincaid Grade Level of 12th grade, indicating it is more difficult to read. In contrast, other units such as Unit 2B scored lower with a CEFR-J level of A1.3, but still presented a higher word difficulty than *FCI*. The verbs per sentence also varied significantly, ranging from 2 to 4.67, adding to the complexity of the texts.

There is an introduction of more abstract and technical vocabulary at earlier stages, which reflects the textbook's goal of preparing students for business and professional environments. For example, Unit 1 includes words such as “accounting” and “advisor,” which are less concrete and more context-dependent. Additionally, the progression toward abstract and professional terminology becomes evident in later units such as Unit 10, which includes words like “log” and “shipment,” both of which require students to comprehend more specialized usage. This shift towards abstract vocabulary poses challenges for lower-level learners, particularly when transitioning from the more concrete language of *FCI*.

TSI also introduces much more complex sentence structures from the very beginning. For example, Unit 1A already has two verbs per sentence (A2.2), and Unit 1B jumps to 4.67 verbs per sentence (C2), making it significantly more complex than anything found in *FCI*. Throughout the textbook, the average number of verbs per sentence remains high, indicating a rapid increase in syntactic complexity. For instance, Unit 2A has 3.75 verbs per sentence

(C2), which requires students to process longer and more intricate sentence structures. This sudden increase in syntactic complexity may create a steep learning curve for students who have been accustomed to the shorter, simpler sentences in *FCI*. The jump from 1-1.5 verbs per sentence in *FCI* to 3-4.67 verbs per sentence in *TSI* could overwhelm students, especially those who struggle with more advanced grammatical constructions such as relative clauses, passive voice, and conditional statements.

CEFR-J Level Consistency

One of the key findings from the analysis is the consistency of difficulty levels across the units:

FCI

The CEFR-J levels are relatively consistent across units in this textbook, ranging from PreA1 to A1.2 for vocabulary and reading. This consistency makes it suitable for use as a foundational textbook for lower-level learners and shows a slight progression of difficulty from unit to unit.

TSI

In contrast to *FCI*, *TSI* shows variability in CEFR-J levels, with some units reaching C1 (e.g., Unit 1B) and others remaining around A1-B1. The variation in levels may lead to challenges for students, particularly if *TSI* is used alongside *FCI*.

Flesch-Kincaid Grade Level

The analysis revealed distinct differences in the readability of the two textbooks.

FCI

The majority of reading passages in *FCI* were found to have Flesch-Kincaid Grade Levels ranging between 6th and 7th grade, indicating that the texts are relatively easy to read for lower-level learners. This aligns well with the textbook's target audience, as students at the A1-A2 levels typically need texts with simple sentence structures and familiar vocabulary. For instance, Unit 1 had a Flesch-Kincaid Grade Level of 7th grade, and Unit 5 had a level of 6th grade, both of which suggest an appropriate level of difficulty for beginner EFL students.

Unit 6 and Unit 12 stood out as exceptions, with the samples of the reading section showing Flesch-Kincaid Grade Levels of 8th/9th grade and 7th grade, respectively. Although these units are still within a manageable range for many learners, the increased difficulty could present challenges for students who are less proficient or who have not yet developed the reading stamina needed to tackle longer and more complex sentences. This suggests that while *FCI* is generally well-suited for beginner learners, instructors should be aware of these outliers and may need to provide additional support or scaffolding to ensure that all students can engage with the material effectively.

TSI

The Flesch-Kincaid Grade Levels in *TSI* were consistently higher than those in *FCI*, with several units reaching advanced levels. For example, the reading section in Unit 1B scored at a 12th-grade level, reflecting a significant increase in sentence complexity and word difficulty. This is particularly concerning for beginners, as texts at this level may require reading proficiency that far exceeds their current abilities. Other units, such as Unit 2B, which had a Flesch-Kincaid Grade Level of 7th grade, still fall within the range of "fairly easy to read" but present a greater challenge than most passages in *FCI*.

The higher grade levels in *TSI* suggest that the textbook is better suited for intermediate or advanced learners. The 12th-grade level texts, in particular, may be too demanding for students who are still developing their basic language skills. These findings raise questions about the appropriateness of using *TSI* alongside *FCI* in beginner-level classes, as the disparity in readability may lead to frustration and decreased motivation among students. Instructors using *TSI* in beginner courses may need to carefully select or adapt readings to ensure that they are accessible to all learners.

Discussion

The results of this study highlight key differences in the vocabulary and readability levels between *FCI* and *TSI*, offering insights into their appropriateness for lower-level EFL learners, particularly in a Japanese university context.

Appropriateness for Beginner-Level Learners

One of the primary findings is the consistent beginner-level readability of *FCI* in terms of both vocabulary and reading sections. With CEFR-J levels ranging from PreA1 to

A2.2 and Flesch-Kincaid Grade Levels from 6th to 7th grade, *FCI* is well-suited for entry-level learners. The stable progression in both vocabulary and readability ensures that students can engage with the material without becoming overwhelmed by difficult texts.

TSI presents a much more varied profile. While some units fall within the A1 to B1 range, other sections contain more advanced vocabulary and higher Flesch-Kincaid scores (up to 12th grade), making the texts significantly more challenging for beginners. This lack of consistency could hinder the development of students' reading skills, as they may face texts that are either too easy or too difficult depending on the unit. The variability also raises concerns about the suitability of using *TSI* alongside *FCI* for lower-level learners. Furthermore, the advanced content in *TSI* could lead to frustration and disengagement among students, as it may not align with their current proficiency level.

The inclusion of higher-level vocabulary and complex sentence structures in *TSI* might require significant supplementary instruction or scaffolding to ensure that students can comprehend and benefit from the material. Instructors may need to modify or adapt *TSI* readings to better align with the students' current proficiency levels or select an alternative textbook with greater consistency and lower readability scores to complement *FCI*. Because *FCI* is taught to lower-level classes, it might be more beneficial to focus solely on *FCI* or introduce a textbook with a more gradual increase in difficulty, ensuring that students' confidence and fluency can build progressively without the added challenge of abrupt changes in text complexity.

The results above seem to echo a number of other studies concerning English textbooks. In a study by Tasaufy (2017), a majority of the textbooks tested were not level-appropriate for students. Moreover, an analysis of Malaysian secondary school textbooks revealed a disproportionate amount of narrative texts that were below the expected grade level (Sidek, 2012). This is hardly surprising. Publishers have the task of creating content for a very broad range of students, and even when a target audience is prescribed based on standardized test scores (e.g., a publisher including a specific learner TOEIC range), such tests are not perfectly reliable (Nicholson, 2015).

Progression

Results indicate that if *FCI* and *TSI* must both be used over the period of a course, special care should be taken in order to prepare students to jump from the former to the latter. At the university where this study was undertaken, courses where these texts are taught are

conducted four to five times per week. Most teachers split the text up in several ways. One common method is to teach material from *FCI* on some days of the week and material from *TSI* on the other days. Some teachers may alternate use of the books week by week. However, in light of this analysis, the best approach might be to introduce *TSI* later in the course. Delaying the use of *TSI* until students have become more comfortable with the language skills covered in *FCI* could reduce the cognitive load early in the semester, as students can focus on gaining fluency and accuracy at their own pace. This, in turn, could lead to greater success when transitioning to more advanced material. A gradual introduction of *TSI*'s more difficult vocabulary might also allow for a smoother progression in vocabulary acquisition and reading comprehension.

By providing students with a clear progression path – starting with more accessible material and gradually working toward more difficult texts – teachers can help maintain students' engagement and motivation throughout the course. One problem with this approach, however, is that students take the TOEIC test at times specified by the university, which do not coincide with the end of the course. In order to deal with this, teachers might focus on *TSI* units that are lower in their CEFR-J and Flesch-Kincaid Reading Grade Level such as Unit 5. Skipping *TSI*'s more challenging units, such as Unit 1, may be another good strategy.

Limitations and Future Research

This study focused primarily on the readability and vocabulary levels of *FCI* and *TSI* using CEFR-J, Coh-Metrix, and Flesch-Kincaid metrics. However, it is important to acknowledge that readability formulas, while useful, do not consider aspects of text complexity such as cultural references or students' prior knowledge, which can also influence comprehension. Neither do they take into account the reader's age or metalinguistic knowledge (Abdollahi-Guilani, 2022). Additionally, the relatively short length of the reading sections may have affected the reliability of the readability assessments, as longer texts typically provide more accurate evaluations. Readability formulae are more reliable when they have a large sample size. Future studies may consider including all the text from the entire unit in order to strengthen their findings.

A notable drawback of numerous readability assessments is their dependence on American educational grade levels for evaluation. This dependency can pose challenges, as the readability levels correlated with specific grades may evolve over time. Chall (1979) highlights the rising apprehension that as the readability of textbooks designated for a

particular grade level improves, students' proficiency in comprehending more complex texts diminishes. Consequently, this situation calls for the development of texts with enhanced readability, which subsequently hampers students' capacity to engage with more challenging materials.

Future research could explore the impact of these textbooks on student outcomes like reading comprehension and fluency, by conducting longitudinal studies in classrooms where both the *Four Corners* and *TOEIC Skills* series are used. Such studies could also examine how students perceive the difficulty of each textbook and whether scaffolding techniques such as pre-teaching vocabulary or providing reading support can mitigate the challenges posed by *TSI*'s advanced content.

Conclusion

This study aimed to compare the vocabulary and reading sections of *FCI* and *TSI* in terms of readability and vocabulary difficulty, with the goal of evaluating their suitability for beginner-level EFL students. Using tools such as the CVLA and Coh-Metrix, the analysis revealed notable differences between the two textbooks.

Results demonstrated that *FCI* is largely consistent with the needs of beginner-level learners, offering vocabulary and reading sections that align well with PreA1 to A2 levels. This consistency makes it a suitable choice for instructors seeking materials that provide a steady progression for low-level students. However, certain units in *FCI*, such as Unit 6 and Unit 12, showed higher difficulty levels, reaching Flesch-Kincaid grade levels of 8th/9th and 7th grade, respectively. While these units are slightly more challenging, they do not detract significantly from the overall appropriateness of the textbook for beginners.

In contrast, *TSI* presented a wider range of difficulty levels, with many units introducing vocabulary and reading passages at B1 to C1 levels. The higher AWD and more complex reading sections, such as Unit 1B (which scored at a 12th-grade level), suggest that *TSI* may be too challenging for lower-level students, particularly when used alongside *FCI*. The variability in *TSI*'s difficulty could result in frustration for learners who are not yet prepared for advanced texts, thus impacting their overall learning experience.

No textbook is perfect in terms of its leveling and progression, and it is certainly up to the teacher to scaffold more difficult content for students. However, instructors should be aware of potentially difficult units or sections within a textbook so that more time can be

devoted to them or they can be skipped or rearranged to present a more linear sequence of difficulty.

References

- Abdollahi-Guilani, M. (2022). Readability index and reading complexity in high school EFL textbooks. *Journal of Foreign Language Research*, 12(2), 37–59.
- Benjamin, R. G. (2012). Reconstructing readability: Recent developments and recommendations in the analysis of text difficulty. *Educational Psychology Review*, 24, 63–88.
- Chall, J. S. (1979). Readability: In search of improvement. *Publishers Weekly*, 216(18), 40–41.
- Crossley, S. A., Allen, D. B., & McNamara, D. S. (2011). Text readability and intuitive simplification: A comparison of readability formulas. *Reading in a Foreign Language*, 23(1), 84–101.
- Dreyer, L. G. (1984). Readability and responsibility. *Journal of Reading*, 27(4), 334–338.
- DuBay, W. (2004). *The principles of readability*. Impact Information.
- Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, 32(3), 221.
- Graesser, A. C., McNamara, D. S., Cai, Z., Conley, M. W., Li, H., & Pennebaker, J. W. (2014). Coh-Metrix measures text characteristics at multiple levels of language and discourse. *The Elementary School Journal*, 115, 210–229. <https://doi.org/10.1086/678293>
- Graham-Marr, A., Anderson, J., Howser, R., & Sato, A. (2018). *TOEIC skills 1* (A. Graham-Marr, A. Sato, & P. Woodfall, Eds.). ABAX.
- Gunning, T. G. (2003). The role of readability in today's classrooms. *Topics in Language Disorders*, 23(3), 175–189.
- Hanifa, R. (2018). EFL published materials: An evaluation of English textbooks for junior high school in Indonesia. *Advances in Language and Literary Studies*, 9, 166–174. <https://doi.org/10.37546/jaltsig.cefr1-1>
- Kane, L., Carthy, J., & Dunnion, J. (2006, April 10-12). *Readability applied to information retrieval* [Conference presentation]. ECIR 2006, London, UK. https://doi.org/10.1007/11735106_56
- McNamara, D. S., & Graesser, A. C. (Eds.). (2012). *Applied natural language processing: Identification, investigation and resolution*. IGI Global. <https://doi.org/10.4018/978-1-60960-741-8.CH011>
- McNamara, D. S., Louwerse, M. M., McCarthy, P. M., & Graesser, A. C. (2010). Coh-Metrix: Capturing linguistic features of cohesion. *Discourse Processes*, 47(4), 292–330. <https://doi.org/10.1080/01638530902959943>

- Nicholson, S. J. (2015). Evaluating the TOEIC® in South Korea: Practicality, reliability and validity. *International Journal of Education*, 7(1), 221–233. <https://doi.org/10.5296/IJE.V7I1.7148>
- Richards, J. C., & Bohlke, D. (2019). *Four corners 1* (2nd ed.). Cambridge University Press.
- Runnels, J. (2013). Preliminary validation of the A1 and A2 sub-levels of the CEFR-J. *Shiken Research Bulletin*, 17(1), 3–10.
- Sidek, H. M. (2012). EFL textbook analysis: A case study. *Language and Literacy*, 14(3), 27–45. <https://doi.org/10.20360/G2HP4J>
- Tasaufy, F. S. (2017). The readability level of the reading texts in English textbook entitled 'bahasa inggris x'. *Edulitics (Education, Literature, and Linguistics) Journal*, 2(2), 62–69.
- Tono, Y. (2019). Coming full circle—from CEFR to CEFR-J and back. *CEFR Journal*, 1, 5–17.
- Uchida, S., & Negishi, M. (2018). Assigning CEFR-J levels to English texts based on textual features. *Asia Pacific Corpus Linguistics Conference Proceedings. (ACPLC) Japan*. 463–467
- Wray, D., & Dahlia, J. (2013). Exploring the readability of assessment tasks: The influence of text and reader factors. *REMIE: Multidisciplinary Journal of Educational Research*, 3(1), 69–95. <https://doi.org/10.4471/REMIE.2013.04>
- Wuttisrisiriporn, N., Vinichevit, N., & Usaha, S. (2020). A situational analysis of EFL textbook selection in Thai public schools. *Indonesian TESOL Journal*, 2(2), 71–80.
- Zamanian, M., & Heydari, P. (2012). Readability of texts: State of the art. *Theory & Practice in Language Studies (TPLS)*, 2(1), 45–53.
- Zhang, L., Liu, Z., & Ni, J. (2013). Feature-based assessment of text readability. *Seventh International Conference on Internet Computing for Engineering and Science (ICICSE)*, 51–54. <https://doi.org/10.1109/ICICSE.2013.18>

英語教育センター紀要 第33号

2025年3月15日発行

編集発行 亜細亜大学英語教育センター

代表者 伊藤 裕子

印刷所 (株) 松井 ビ・テ・オ・印刷

発行所 亜細亜大学英語教育センター
〒180-8629

東京都武蔵野市境 5-8

☎0422-54-3111(ダイヤルイン) 内線2475
