

JACET 授業学ジャーナル
JACET Journal of Developmental Education
Vol. 5

March 2025

JACET SIG 授業学研究会（関東・中部・関西）

目次

JACET 授業学ジャーナル 第5号

< 研究論文 >

Practicing Notetaking While Listening to Academic Lectures: EFL Learners' Perceptions and Their Effects on TOEFL ITP® Listening Scores

SAKURAI, Shizuka

SPRING, Ryan.....1

Instructor Perceptions of Their Role in Student Motivation

MOROI, Takako.....20

< 研究ノート >

Structuring Polysemous Words' Instruction: A Hierarchical Approach

IJIMA, Hisanori.....37

< 実践報告 >

How ChatGPT Improves Writing for EFL Students

MATSUNO Sumie.....56

教員養成課程の英語授業における英語スピーキング能力向上のための多様な学習活動とその効果

瀧澤 典子.....67

DeepL Write を活用した英語ライティング指導と訂正フィードバック

甲斐 順.....87

大学英語教育における ChatGPT の導入：ライティング学習支援効果と課題の検証

古藪 麻里子

藤城 晴佳

町村 貴子.....107

専門留学における期待・不安要素について-保育専攻の大学生に対する
事前事後の調査から-

加藤 あや美

内田 政一

ランズベリー ローレン

...124

Table of Contents

JACET Journal of Developmental Education Vol. 5

<Research Paper>

- Practicing Notetaking While Listening to Academic Lectures: EFL Learners' Perceptions and Their Effects on TOEFL ITP® Listening Scores
SAKURAI, Shizuka
SPRING, Ryan.....1

- Instructor Perceptions of Their Role in Student Motivation
MOROI, Takako.....20

<Research Note>

- Structuring Polysemous Words' Instruction: A Hierarchical Approach
IIJIMA, Hisanori.....37

<Case Study>

- How ChatGPT Improves Writing for EFL Students
MATSUNO Sumie.....56

- The effect of varied activities for the development of EFL speaking performance among students of a teacher certification program
TAKIZAWA Noriko.....67

- Instruction and Feedback in English Writing Using DeepL Write
KAI, Jun.....87

- Implementation of ChatGPT in University English Education: Its Effects and Challenges in Writing Skill Development
FURUYABU, Mariko
FUJISHIRO, Haruka
MACHIMURA, Takako....107

Expectations and Anxieties Regarding A Specialized Study Abroad Program:
A Pre-and Post-Survey of University Students

KATO, Ayami

UCHIDA, Masaichi

LANDSBERRY, Lauren...124

What Affects EGAP Lecture Listening? The Dynamics of Practice, Notetaking Ability, and Perceived Difficulty in an EFL Context

Shizuka Sakurai (Tohoku University)

Ryan Spring (Tohoku University)

Keywords: lecture listening, notetaking, lecture listening, EFL, EGAP, perceived difficulty

Abstract

This study explores the impact of teaching notetaking to Japanese EFL learners on listening comprehension. It addresses conflicting findings in the literature, with some research suggesting that notetaking can hinder comprehension, while others highlight its benefits. A total of 133 participants practiced notetaking strategies while listening to academic English lectures over the course of a semester. The findings indicate that students' notetaking skills significantly improved, as evidenced by an increase in the number of words written and a decrease in perceived difficulty. Multiple regression analyses revealed that listening comprehension scores and feelings of task difficulty were predictors of EGAP listening performance. However, the number of words in notes was not strongly correlated with immediate EGAP listening ability, although it may contribute to long-term improvement. These findings highlight the importance of addressing both practical notetaking skills and subjective perceptions when teaching listening strategies in EFL contexts.

Introduction

Notetaking is an important skill in academic contexts for both first (L1) and second (L2) languages (Jin & Webb, 2023; Salame et al., 2024; Voyer et al., 2022). Notes act as external memory, aiding in review, exam preparation, and other tasks (Kiewra et al., 1995; Laidlaw et al., 1993; Siegel, 2021). The act of notetaking also enhances retention, attention, and comprehension (Carrier, 1983), leading to significant educational benefits (Carrell et al., 2002). For EFL (English as a Foreign Language) learners, notetaking while listening to a lecture simulates the university classroom experience where English is the medium

(Hansen, 1994; Sakurai, 2018; Siegel, 2021, 2022). Consequently, it is included in many EGAP (English for General Academic Purposes) curricula.

Despite its importance for academic performance in English-medium classes (e.g., Siegel, 2021), Japanese students often lack training in notetaking during English lectures prior to university (Crawford et al., 2016; Sakurai, 2018). As a result, university students are not familiar with the task of notetaking or the strategies involved. This study introduces how these skills were instructed and practiced over the course of a semester in EFL classes at a Japanese university. It also examines the effects on learners' listening, notetaking, and perceived difficulty, and explores how these factors are associated with learners' EGAP ability, as measured by the TOEFL ITP® test.

Literature Review

Studies examining the effects of notetaking on L2 lecture comprehension have shown inconsistent results. Some studies suggest that notetaking has negative effects on improving L2 listening ability. For example, Clark et al. (2013) and Hale and Courtney (1994) found that notetaking did not promote TOEFL listening section scores, but rather, impeded students' comprehension. Furthermore, Chaudron et al. (1994) reported that the availability of notes (for TOEFL listening parts) does not have an impact on the result of listening on immediate recall tests. Nekoda (2020) found that EFL learners' listening-comprehension scores improved after notetaking practice but declined by a final test, indicating reduced effectiveness over time. On the other hand, Jin and Webb (2023) conducted a meta-analysis of 21 studies on the effectiveness of notetaking and found that learners who were instructed to take notes tended to show higher learning outcomes than those who were allowed to choose whether or not to take notes. Similarly, positive effects of EFL notetaking training over a particular period (a semester or year) have been reported, specifically: large gains in the total number of notations, content words, abbreviations, arrows, and in listening comprehension (Crawford, 2015; Jin & Webb, 2023; Sakurai, 2018; Siegel, 2015).

One reason for the seemingly conflicting results presented above may be the learning and testing conditions. Researchers have investigated some important factors affecting listening and notetaking performance (Bloomfield et al., 2010; Carrell et al., 2002; Sakurai, 2024; Song, 2012). They claimed that ESL/EFL students with “better” notes tend to perform well in listening, indicating that instruction on taking effective notes may promote listening. Better notes are generally defined as those that include key points and

important detailed information while maintaining an organized overall structure (Song, 2012). Carrell et al. (2002) found that learners taking notes in L2 performed better in shorter lectures (2 minutes vs. 5 minutes). Similarly, the speaking rate of the lecturer also affects notetaking, with faster speech making it harder to take notes (Piolat et al., 2004; Sakurai, 2018). Additionally, complex listening materials with unfamiliar words, topics, and syntax increase difficulty (e.g. Crossley et al., 2008; Jeon & Yamashita, 2014; Sakurai & Spring, 2022). Factors such as these might also make the learner's proficiency level important, as a mismatch between the learner level and materials level can impede acquisition and uptake (e.g., Tomlinson, 2013). Unfortunately, most studies on L2 notetaking don't control for these linguistic features, limiting conclusions about its impact on listening comprehension. However, in actual classrooms, listening materials are generally diverse, making it difficult and unrealistic to control all variables in the same way.

Research Questions

Based on the findings of previous studies, there is still uncertainty regarding whether students' notetaking abilities improve over time with consistent practice. Furthermore, it is unclear how certain aspects of notetaking are linked to EGAP listening ability. To address these issues, this study will examine the effects of repeated practice with lectures carefully selected to ensure a balanced consideration of length, speed, and linguistic difficulty, to avoid bias from these factors. Furthermore, it will attempt to look at factors of notetaking (i.e., the number of words written, comprehension, and student perceived difficulty) and consider this data with TOEFL ITP® test listening section scores to see which factors are important for listening skills in general, and which are important for improvement in EGAP listening. Specifically, we seek to address the following two research questions:

RQ 1. Does students' ability to take notes while listening improve over time with practice?

RQ 2. What aspects of notetaking are associated with better EGAP listening ability and its development?

Method

Participants

A total of 173 first-year EFL students from various departments at Tohoku University practiced

listening to lectures and taking notes. The classes consisted mainly of Japanese students, with four international students excluded from the dataset due to their different educational backgrounds and English proficiency levels. Of the remaining students, only the data from 133 who completed all tasks and provided ethical consent were analyzed. The majority of these participants ($n = 112$) who took the TOEFL ITP® test in May were classified as B1 on the CEFR scale, indicating an intermediate level of English proficiency. The remaining 21 students achieved either B2 or A2 levels. The students were enrolled in a required EGAP course, “Academic Listening and Speaking,” and there were about 43 students per class: two classes for engineering majors, one for nursing and dental majors, and one for humanity majors. In these classes, taught by the first author, students met once a week for 90 minutes and were expected to learn and practice speaking and listening skills, particularly notetaking while listening, orally summarizing from notes, interrogative and giving opinions, and idiomatic expressions. The participants also took another required EGAP course, “Academic Reading and Writing,” which also met once a week but was taught by different teachers. All participants received guidance on academic lectures and notetaking in English at the beginning of the semester and regularly practiced the skills and strategies over a semester. Our study received ethical approval from the university’s research ethics committee. Ethical considerations were ensured for all participants, including informed consent, confidentiality, and the voluntary nature of participation.

Notetaking Instructions and Practice

Instruction on academic notetaking was provided before students engaged in actual English lecture listening activities. At the beginning of the semester, students read textbook explanations on notetaking while listening, in accordance with the school’s mandated curriculum, and received guidance on the following topics in class:

1. The importance and role of notetaking (such as information acquisition and retention, active listening, and focus).
2. How to listen and take notes (paying attention to key information, learning and identifying signal words and phrases, and organizing and structuring information).
3. Techniques for quick notetaking (using abbreviations and symbols)
4. Preparatory actions (previewing materials before class, such as reading documents).

5. Listening to lectures while reviewing sample notes, and grasping notetaking tips.
6. Practicing notetaking and sharing and reinforcing information in groups.

Though these points were mandated by the students' curriculum, we feel that point three is particularly important. One reason that Japanese university students have difficulty with the notetaking while listening task is because of the differences in the orthographic systems and syntax in Japanese and English. However, studies such as Crawford (2015), Sakurai (2018), and Siegel (2015) suggest that the use of symbols and abbreviations may mitigate the effect of these differences and underscore the importance of notetaking training for Japanese EFL learners. In particular, since the learners were not familiar with symbols and abbreviations, a list of these was provided (see Appendix A). Learners filled in the list of abbreviations and symbols at home. Students practiced using abbreviations and symbols through various activities, such as filling in lists at home, pair-work, and comparing notes in groups to discuss efficient notetaking strategies. They also listened to lectures while referring to sample notes, which helped them understand quick and effective ways to capture and organize information.

After receiving the instruction, students listened to English-language lectures weekly throughout the semester (15 weeks) in the classroom, focusing on various academic topics, and took notes in English during class. Each lecture was listened to only once, and afterward, the students had to complete a 10-question comprehension quiz written in English. The quality of the quiz questions was ensured by having several reviews amongst the authors and then checking the internal reliability of the questions via Cronbach's alpha per quiz (see Table 1) and the Cronbach's alpha score of students' overall scores across the nine lectures; $\alpha = 0.42$. Students took a photo of their notes after each session and submitted it via Google Forms, along with a report detailing the number of words, abbreviations, and symbols they used. They were allowed to reference their notes during the quiz. The results from these quizzes provided insights into their listening comprehension skills and notetaking skills.

We selected some longer and shorter lectures in length. We measured the length of a lecture by the number of words spoken in the lecture and the duration (in minutes and seconds) of the lecture audio. We measured speech rate in words per minute, and complexity via the Flesch-Kincaid Grade-level metric (FKG) because it is generally tied to comprehensibility (e.g. Kincaid et al., 1975), but also because it is calculated in part based on the number of syllables in given words. They tend to impact the difficulty in

listening to lectures, specifically, as higher FKG scores would indicate more multi-syllable words, which would theoretically be more difficult to comprehend for learners. Table 1 presents the details of the lectures used, including the topic, duration, word count, and FKG.

Table 1

List of Lectures for Notetaking While Listening Activities and Characteristics

	Duration	Word Count	α of Comp. Questions	Speech Rate (WPM)	Readability (FKG)	Topic
Lecture 1	3:22	474	0.13	141	11.5	College Orientation
Lecture 2	3:16	407	0.34	125	11.9	Diet and behavior
Lecture 3	7:15	807	0.16	113	7.1	Happiness
Lecture 4	6:00	837	0.26	140	6.0	Sleep
Lecture 5	3:39	441	0.26	121	10.5	Sea otters
Lecture 6	3:37	431	0.21	119	11.1	Ozone depletion
Lecture 7	6:27	768	0.20	119	7.7	Video Games
Lecture 8	6:00	821	0.29	137	7.5	Technology: Robots
Lecture 9	3:16	407	0.34	125	11.9	Diet and behavior

In total, the students practiced lecture listening and notetaking nine times during the semester (Table 1). There were two types of lectures based on length: 4 shorter lectures (approximately 3 minutes, Lectures 1, 2, 5, and 6) and 4 longer lectures (approximately 6 to 7 minutes, Lectures 3, 4, 7, and 8). The duration and word count of these lectures were about double that of the shorter lectures. The same lecture was used as a pre-test and post-test (Lectures 2 and 9) to avoid the influence of various factors in the lectures that could affect learners' listening abilities and notetaking skills differently. Some lectures were originally written for the English courses of the university, and some were from the collections of lectures in the EFL books, *Instant Academic Skills* (Lane, 2011) and *Contemporary Topics I* (Solórzano & Frazier, 2009). The multiple-choice questions for comprehension quizzes in English were created by the authors of the present study. They consisted of 10 questions in Google Forms that asked about the lecture's purpose, main points, key details, and data (e.g., Questions: “1. What is the main topic of this lecture?”, Choices: a) The impact

of technology, b) Language and the universe on language, c) Theories about the origin of language and its evolution, d) How our ideas of language changed over time). At home, the students were also assigned to listen to different short lectures using the textbook, *Official Guide to the TOEFL ITP® Assessment Series* (Educational Testing Service, 2019), and other in-house materials.

After answering the comprehension questions, students were instructed to count the number of words in their notes and respond to a post-lecture survey question about the perceived difficulty of the task: "Was notetaking while listening difficult?" on a Likert scale from 1 to 5. The word counts were reported by the participants and verified by research assistants. Since most students took notes on paper, with only a few (about 3 to 5 students per class) using electronic devices, we did not examine any differences between digital and paper-based notetaking methods.

Although there are some merits of using multilingual resources, L1 or both L1 and L2, in notetaking (Koren, 1997), this approach is more effective for students with lower proficiency levels and not challenging enough for students with higher levels (Siegel, 2022). Therefore, in the present study, we advised the learners (mostly CEFR B1 level) to take notes in L2 (English), not in L1 (Japanese). However, a few words in the L1 were allowed at the beginning, hoping that students would gradually rely less on their L1 when notetaking. Students practiced summarizing in English using their notes, with ongoing reminders to take notes in English. Notetaking skills were evaluated based on note quantity, quality, and organization of key information in the final test.

Analysis

To check for any discernible differences before and after the treatment, we utilized the scores from lectures 2 and 9, as they were the same lectures and questions. The lecture listening comprehension scores and number of words written were generally considered continual in nature, whereas the perceived difficulty was considered ordinal data due to its being measured on a Likert scale. Dependent *t*-tests were used for the measures due to the continuous and normal nature of the data sets as verified by a Shapiro-Wilk test of normality, and a Wilcoxon signed-rank test was used for the self-reported difficulty due to the ordinal nature of the data (Spring, 2022). Cohen's *d* was reported as a measure of effect size for *t*-tests and Spearman's *rho* was reported for the effect size of Wilcoxon-signed rank tests, and these were interpreted according to Plonsky and Oswald (2014).

To determine how well the number of notes, perceived difficulty of listening and notetaking, and lecture comprehensibility correspond to both EGAP listening ability and changes in EGAP listening ability, we conducted two multiple regression analyses to predict students' TOEFL ITP® Section 1 (Listening Comprehension) scores. We created two models, one to predict TOEFL ITP® scores that the learners took in May to investigate whether these measures were related to general EGAP ability, and another one using the delta scores (post-test minus pre-test) of another TOEFL ITP® test the students took in November, four months after the treatment had concluded, and the original test taken in May. Ideally, TOEFL pre- and post-tests would have been given in April and August, but since the schedule was decided by the university, we used the available data: the scores that were taken in May and November. In April, orientation and TOEFL explanations were given, followed by notetaking and lecture listening training. Since learners had not yet received much training when the first TOEFL test was taken in early May, we believe that it is still valid. From April to August, learners practiced notetaking while listening to lectures, and in the fall, they should have received more practice from their second-semester teachers. This led to skill refinement, reflected in the November TOEFL test, which was considered a valid post-test.

For predictor variables in the first model, we used the sum of students' comprehension scores on the lecture listening activities, the sum of the number of words they wrote on all activities, and the sum of their self-reported difficulty with the lectures. For the second model, we used the May TOEFL ITP® score to account for previous ability, and the delta scores of the same three variables from the first model to ascertain whether improvement in these skills correlated with improvement in EGAP listening ability. We used dominance analysis to determine the relative weight of each predictor variable to help us interpret which had the association with the dependent variables, and random forests to determine whether or not each dependent variable was significantly associated with the predictor variable following Mizumoto (2023).

Results

The descriptive data on students' performance in each lecture and the delta values (Lecture 9 minus Lecture 2) are presented in Table 2. There was considerable variability in the number of words in their notes, with moderate variability in comprehension scores and perceived lecture difficulty. Additionally, the differing lengths of the lectures contributed to variations in note length.

Table 2

Descriptive Statistics for Student Performance on Lecture Listening Activities

Lec.	Comprehension score	No. words written	Self-reported difficulty
1	1–10; $M = 6.35$; $S = 1.44$	2–63; $M = 30.8$; $S = 10.8$	3–5; $M = 4.60$; $S = .55$
2	2–9; $M = 6.03$; $S = 1.47$	12–64; $M = 32.6$; $S = 10.9$	2–5; $M = 4.68$; $S = .54$
3	3–10; $M = 7.03$; $S = 1.50$	36–210; $M = 72.8$; $S = 23.9$	2–5; $M = 3.95$; $S = .77$
4	3–10; $M = 7.78$; $S = 1.30$	10–138; $M = 68.5$; $S = 20.4$	2–5; $M = 4.01$; $S = .74$
5	3–10; $M = 7.42$; $S = 1.39$	13–75; $M = 38.5$; $S = 12.7$	2–5; $M = 3.74$; $S = .75$
6	3–9; $M = 6.25$; $S = 1.41$	8–96; $M = 44.7$; $S = 14.5$	2–5; $M = 3.90$; $S = .76$
7	3–10; $M = 7.92$; $S = 1.27$	46–250; $M = 86.7$; $S = 25.5$	2–5; $M = 3.33$; $S = .83$
8	3–10; $M = 6.43$; $S = 1.50$	31–136; $M = 70.2$; $S = 19.2$	2–5; $M = 4.22$; $S = .80$
9	3–10; $M = 7.44$; $S = 1.51$	20–101; $M = 50.3$; $S = 12.8$	1–5; $M = 3.95$; $S = .89$
Delta	-3–6; $M = 1.41$; $S = 1.85$	-12–51; $M = 17.7$; $S = 10.9$	-4–2; $M = -.74$; $S = .94$

Note. Lec. = Lecture. Range is provided, then M = mean, S = standard deviation

The results of the comparative pre-test and post-test measures are provided in Table 3. They show that there were statistically significant differences in all measures and that there was a large positive effect on the number of words that students wrote in their notes, a large negative effect on perceived difficulty, and a medium-sized positive effect on listening comprehension scores. This indicates that through the treatment, students in this study, on average, improved their listening ability a modest amount, and became much more able to take notes (as shown in Appendix B), perceiving the task as far less difficult.

Table 3

Pre- and Post-test Difference Testing Statistics

Comprehension Scores	$t(132) = 8.78, p < .01, d = 0.76^{**}$
Number of Words Written in Notes	$t(132) = 18.75, p < .01, d = 1.63^{***}$
Self-Reported Difficulty	$Z = 7.04, p < .01, r_s = 0.79^{***}$

Note. ***large effect size; **medium effect size

The results of the multiple regression analysis predicting May TOEFL ITP® listening scores, presented in Table 4, explained 21.6% of the variance; $R^2 = .216$, $F = 13.14$, $p < .001$. Students' scores on the listening comprehension questions explained the largest part of the model (66%, Table 4), and both these scores and their subjective feelings of difficulty with the activity (29%, Table 4), were significant predictors. The results suggest that the number of words in students' notes was not strongly correlated with their EGAP listening scores, but that students who perceived the task of notetaking while listening as difficult were likely to have lower EGAP listening ability, as reflected in their TOEFL ITP® listening scores. These findings highlight the importance of addressing subjective factors such as task difficulty perception when considering ways to improve listening comprehension and overall EGAP performance.

Table 4

Regression Analysis for May TOEFL ITP® Listening Scores and Average Lectures Scores (1-9)

Variable	B	SE	β	t	p	R.W.	R.F.
Listening Comp.	0.22	0.05	0.38	4.76	0.00	0.16 (66%)	Confirmed
No. Words / Notes	0.00	0.00	0.02	0.19	0.85	0.01 (5%)	Rejected
Self-reported Diff.	-0.16	0.06	-0.23	-2.65	0.01	0.07 (29%)	Confirmed

Note. R.W. = Relative weight. R.F. = Random Forest

The results of the multiple regression analysis predicting changes in TOEFL ITP® listening scores, presented in Table 5, explained 21.4% of the variance; $R^2 = .214$, $F = 8.77$, $p < .001$. The largest predictor was the initial TOEFL ITP® scores (74%), which is to be expected. However, with these scores included in the model, a decrease in self-reported difficulty was still found to be a significant predictor of improvement, indicating that overcoming the difficulty of this task likely played some role in EGAP listening improvement.

Table 5

Regression Analysis for TOEFL ITP® Delta Listening Scores and Lecture Delta Scores

Variable	B	SE	β	t	p	R.W.	R.F.
May TOEFL Scores	-0.42	0.08	-0.40	-5.03	0.00	0.16 (74.2%)	Confirmed
Listening Comp.	0.16	0.14	0.09	1.18	0.24	0.01 (3.5%)	Rejected
No. Words / Notes	0.04	0.02	.131	1.667	0.09	0.02 (8.6%)	Rejected
Self-reported Diff.	-0.64	0.27	-.185	-2.35	0.02	0.03 (13.7%)	Confirmed

Note. R.W. = Relative weight. R.F. = Random Forest

Discussion

This study aimed to explore the effectiveness of notetaking instruction and practice on Japanese university students' listening and notetaking abilities over time, specifically focusing on two research questions: whether students' ability to take notes while listening improved with practice and which aspects of notetaking were associated with better EGAP listening ability and improvements in listening ability, as measured by the TOEFL ITP®.

Improvement in Notetaking Ability

The results indicate that students could improve their notetaking skills over the semester. This was evidenced by an increase in listening comprehension scores, the number of words taken in notes, and a decrease in perceived difficulty from the first to final lectures. This aligns with previous studies that emphasize the necessity of notetaking training for Japanese EFL learners. The structured practice sessions, including instruction on effective strategies for listening and notetaking, likely helped students to gradually develop their skills. Additionally, the introduction of common abbreviations and symbols, despite initial unfamiliarity, likely contributed to students' ability to take more comprehensive notes, as evidenced by the increase in average word counts. Furthermore, reducing cognitive load through improved notetaking may enhance listening ability by allowing better resource allocation, increasing note quantity, improving comprehension, and lowering perceived difficulty.

Association Between Notetaking and EGAP Listening Ability

Regarding the second research question, the multiple regression analyses showed that students' lecture listening comprehension and perceived difficulty in notetaking while listening were associated with EGAP listening. Moreover, decreased perceptions of task difficulty were associated with long-term improvement in EGAP listening ability. These findings suggest that the ability to take notes while listening to English lectures is closely linked to EGAP listening performance. This finding highlights the subjective nature of task difficulty and its potential impact on performance. Addressing students' perceptions of difficulty in notetaking tasks may be essential for enhancing their overall listening comprehension and academic performance. It can also be explained that the practice of organizing notes using abbreviations, symbols, lists, and indentations, as taught in the instruction, may have contributed to reducing perceived difficulty and improving listening comprehension.

One reason that the number of words written in notes was not associated with EGAP listening performance could be due in part to the fact that high-quality notes do not necessarily contain more words. However, our current dataset does not have a measure of quality to explore this further. Furthermore, the reason that the listening comprehension scores did not better predict TOEFL ITP® listening improvement scores could be due in part to the quality of the questions. Specifically, the Cronbach's alpha scores for comprehension questions were rather low, so it could be that some of the questions were not truly representative of L2 listening skill. Future research should focus on improving the reliability and validity of listening comprehension assessments, possibly by using more precise and consistent question formats.

Finally, it is important to note that the correlation coefficients and the percentage of variance explained in this study were not particularly high. This may be because the TOEFL ITP® listening section has 50 questions, but only 12 of them (24%) are lecture-based. As a result, notetaking would be expected to account for only about 24% of the variance in these scores.

Implications for Instruction

The findings suggest several implications for instructional practices in EFL settings. First, providing explicit notetaking training, including familiarization with abbreviations and strategies for identifying key information, can significantly enhance students' notetaking skills. Second, educators should address the subjective factors that influence students' experiences, such as their perceived difficulty with notetaking

tasks. Incorporating regular feedback and supportive learning environments can help mitigate students' feelings of difficulty, ultimately leading to better academic outcomes. Lastly, varying lecture lengths and complexities in practice activities could enhance students' adaptability and effectiveness in real lecture environments.

Challenges and Future Research

Despite some interesting findings, this study faces several challenges that require further exploration. One limitation is the reliance on self-reported data to measure perceived difficulty in notetaking. Self-reports can be subjective and may not reflect the real challenges students face due to individual perception differences. Additionally, the broad nature of the Likert scale used may not capture the nuances of each student's experience. Additionally, the Cronbach's alpha scores for the lecture comprehension questions indicate that some caution is needed when interpreting each lecture's listening comprehension scores. Future studies should utilize improved questions and replicate some aspects of this study.

Next, while this study used a rather rudimentary metric for notes (i.e., the number of words written), the quality of the notes was not taken into account. However, as notes are meant to be concise, we believe that quality is a critical aspect of notetaking that should be investigated in the future. Studies like these could explore how factors such as the content, accuracy, and organization of notes affect students' listening and comprehension. Understanding these dimensions could provide valuable insights for refining notetaking strategies and enhancing academic success among EFL learners.

Lastly, with the advancement of technology, notetaking habits and learning methods have significantly changed (Morehead et al, 2019). Future research should investigate the relationship between notetaking habits, styles, and learning outcomes. Various aspects need to be explored for effective classroom instruction, including notetaking during in-person classes, notetaking in online courses, methods of reviewing notes, and instruction on effective notetaking strategies. This will contribute to more diverse and effective teaching practices.

Conclusion

This study contributes to our understanding of the dynamics between notetaking, listening comprehension, and EGAP performance among Japanese university students. It showed that although notetaking while

listening to an L2 lecture can be challenging for EFL students, repeated practice over time can lead to improvement in terms of listening comprehension, the number of notes written, and perceived difficulty. Furthermore, the findings highlight the significance of notetaking while listening for EGAP listening. Both comprehension scores and perceived difficulty were associated with TOEFL ITP® scores during the treatment, while reductions in perceived difficulty were linked to subsequent improvement. While we were unable to observe the impacts of note quality and should control the quality of listening comprehension questions in the future, we believe these results provide support for the notion that notetaking while listening to lectures should be encouraged and repeatedly practiced in EGAP courses.

Note:

This study was presented by the same authors at the 62nd JACET International Convention and was supported by a JSPS Grant-in-Aid (Category C), Project Number 23K00741.

References

- Bloomfield, A., Wayland, S. C., Rhoades, E., Blodgett, A., Linck, J., & Ross, S. (2010). What makes listening difficult? *University of Maryland Center for Advanced Study of Language*.
<https://apps.dtic.mil/sti/pdfs/ADA550176.pdf>
- Carrell, P. L., Dunkel, P., & Mollaun, P. (2002). The effects of notetaking, lecture length and topic on the listening component of the TOEFL 2000. *TOEFL® Monograph Series* (MS-23).
<https://typeset.io/pdf/the-effects-of-notetaking-lecture-length-and-topic-on-the-35e8wch8bf.pdf>
- Carrier, C. A. (1983). notetaking research: Implications for the classroom. *Journal of Instructional Development*, 6(3), 19–26. <https://doi.org/10.1007/BF02906208>
- Chaudron, C., Loschky, L., & Cook, J. (1994). Second language listening comprehension and lecture notetaking. In J. Flowerdew (Ed.), *Academic listening: Research perspectives* (pp. 75–92). Cambridge University Press.
- Clark, M., Wayland, S., Osthus, P., Brown, K. G., Castle, S., & Ralph, A. (2013). The effects of note taking on foreign language listening comprehension: An empirical study. *University of Maryland Center for Advanced Study of Language*. <https://www.govtilr.org/Publications/Notetaking.pdf>
- Crawford, M. J. (2015). A study on note taking in EFL listening instruction. In P. Clements, A. Krause, &

- H. Brown (Eds.), *JALT2014 Conference Proceedings* (pp. 416–424).
- Crawford, M. J., Ducker, N., MacGregor, L., Kojima, S., & Siegel, J. (2016). Perspectives on note taking in EFL listening. In P. Clements, A. Krause, & H. Brown (Eds.), *Focus on the learner*. JALT. <https://jalt-publications.org/node/4/articles/5399-perspectives-notetaking-efl-listening>
- Crossley, S. A., Greenfield, J., & McNamara, D. S. (2008). Assessing text readability using cognitively based indices. *TESOL Quarterly*, 42(3), 475–493. <https://doi.org/10.1002/j.1545-7249.2008.tb00142.x>
- Educational Testing Service. (2019). *Official Guide to the TOEFL ITP® Assessment Series*. Educational Testing Service.
- Hale, G. A., & Courtney, R. (1994). The effects of notetaking on listening comprehension in the test of English as a foreign language. *Language Testing*, 11(1), 29–47. <https://doi.org/10.1177/026553229401100104>
- Hansen, C. (1994). Topic identification in lecture discourse. In J. Flowerdew (Ed.), *Academic listening: Research perspectives* (pp. 131-145). Cambridge University Press.
- Jeon, E. H., & Yamashita, J. (2014). L2 reading comprehension and its correlates: A meta-analysis. *Language Learning*, 64(1), 160–212. <https://doi.org/10.1111/lang.12034>
- Jin, Z., & Webb, S. (2023). The effectiveness of note taking through exposure to L2 input: A meta-analysis. *Studies in Second Language Acquisition*, 1–23.
- Kiewra, K. A., Benton, S. L., Kim, S., Risch, N., & Christensen, M. (1995). Effects of notetaking format and study technique on recall and relational performance. *Contemporary Educational Psychology*, 20(2), 172–187. <https://doi.org/10.1006/ceps.1995.1011>
- Kincaid, P. J., Fishburne, R. P., Rogers, R. L., & Chissom, B. S. (1975). *Derivation of new readability formulas (Automated readability index, fog count and Flesch reading ease formula) for navy enlisted personnel*. Institute for Simulation and Training, University of Central Florida. <http://www.dtic.mil/dtic/tr/fulltext/u2/a006655.pdf>
- Koren, S. (1997). Listening to Lectures in L2; Taking Notes in L1. *TESOL-EJ*, 2 (4). <http://www.tesl-ej.org/wordpress/issues/volume2/ej08/ej08a1/>
- Laidlaw, E. N., Skok, R. L., & McLaughlin, T. F. (1993). The effects of notetaking and self-questioning on quiz performance. *Science Education*, 77(1), 75–82. <https://doi.org/10.1002/sce.3730770105>
- Lane, S. (2011). *Instant Academic Skills*, Cambridge University Press.

- Mizumoto, A. (2023). Calculating the relative importance of multiple regression predictor variables using dominance analysis and random forests. *Language Learning*, 73(1) 161–196.
<https://doi.org/10.1111/lang.12518>
- Morehead, K., Dunlosky, J., Rawson, K. A., Blasiman, R., & Hollis, R. B. (2019). Notetaking habits of 21st Century college students: implications for student learning, memory, and achievement. *Memory*, 27(6), 807–819. <https://doi.org/10.1080/09658211.2019.1569694>
- Nekoda, H. (2020). L1 oyobi L2 ni yoru nōtoteikingu shidō: Keizoku shidō ni yoru kōka no kenshō [Note-taking instruction using L1 and L2 lectures]. *CASELE JOURNAL Chubu*, 50, 23–35.
https://doi.org/10.18983/caselejournal.50.0_23
- Piolat, A., Olive, T., & Kellogg, T. R. (2004). Cognitive effort during note taking. *Applied Cognitive Psychology*, 19(3), 291–312. <https://doi.org/10.1002/acp.1086>
- Plonsky, L., & Oswald, F. L. (2014). How big is “big”? Interpreting effect sizes in L2 research. *Language Learning*, 64(4), 878–912. <https://doi.org/10.1111/lang.12079>
- Sakurai, S. (2018). Promoting skills and strategies of lecture listening and note-taking, *Fukuoka University Review of Literature & Humanities* 49(4). 1019-1046.
- Sakurai, S., & Spring, R. (2022). Analysis of difficulty in listening to lectures: What makes lectures more or less difficult for Japanese EFL students? *Annual Report of the Center for Language and Culture Education, Institute for Excellence in Higher Education, Tohoku University*, 8, 26–30.
- Sakurai, S. (2024). Investigating the production and evaluation of academic English listening materials with generative AI. *Bulletin of the Institute for Excellence in Higher Education, Tohoku University*, 10, 97–104.
- Salame, I. I., Tuba, M., & Nujhat, M. (2024). Note-taking and its impact on learning, academic performance, and memory. *International Journal of Instruction*, 17(3). <https://e-iji.net/ats/index.php/pub/article/view/630>
- Siegel, J. (2015). A pedagogic cycle for EFL notetaking. *ELT Journal*, 70 (3), 225–227.
<https://doi.org/10.1093/elt/ccv073>
- Siegel, J. (2021). *Developing Notetaking Skills in a Second Language, Insights from Classroom Research*. Routledge.
- Siegel, J. (2022). Factors affecting notetaking performance, *International Journal of Listening*, 1–

13. <https://doi.org/10.1080/10904018.2022.2059484>
- Solórzano, H., & Frazier, L. (2009). *Contemporary Topics 1*. Pearson-Longman.
- Song, M. (2012). notetaking quality and performance on an L2 academic listening test. *Language Testing*, 29 (1), 67–89. <https://doi.org/10.1177/0265532211415379>
- Spring, R. (2022). Free, Online Multilingual Statistics for Linguistics and Language Education Researchers. *Center for Culture and Language Education, Tohoku University 2021 Nenpo*, 8, 32–38. <https://doi.org/10.13140/RG.2.2.12037.63202>
- Tomlinson, B. (2013). Second language acquisition and materials development. In B. Tomlinson (ed.), *Applied Linguistics and Materials Development* (pp. 11–33). CUP. <https://doi.org/10.1017/9781139042789>
- Voyer, D., Ronis, S. T., & Byers, N. (2022). The effect of notetaking method on academic performance: A systematic review and meta-analysis. *Contemporary Educational Psychology*, 68. <https://doi.org/10.1016/j.cedpsych.2021.102025>

Appendices

Appendix A: Practice with Abbreviations and Symbols

	ABBREV. / SYMBOL	Meaning		ABBREV. / SYMBOL	Meaning
1	=		1	ppl.	
2	≠		2	dept.	
3	≈		3	govt.	
4	&, +		4	nat'l.	
5	>		5	int'l	
6	<		6	impt.	
7	∴		7	b/w	
8	∴		8	ch.	
9	→		9	esp.	
10	←		10	excl. / incl.	
11	↑		11	def.	
12	↓		12	diff.	

(Sample worksheet)

	ABBREV. / SYMBOL	Meaning		ABBREV. / SYMBOL	Meaning
1	=	is equal to, is the same as	1	ppl.	people
2	≠	differ from	2	dept.	department
3	≈	is similar to	3	govt.	government
4	&, +	and, in addition to	4	nat'l.	national
5	>	greater than, more than	5	int'l	international
6	<	less than, smaller/fewer than	6	impt.	important
7	∴	because	7	b/w	between
8	∴	therefore, thus	8	ch.	chapter
9	→	lead to, result in, affect	9	esp.	especially
10	←	result from, be affected by	10	excl. / incl.	excluding, including
11	↑	go up, increase, rise	11	def.	definition
12	↓	go down, decrease, drop	12	diff.	different, difference

(Sample worksheet with answers)

Note: When taking academic lecture notes, university students commonly use various abbreviations to save time and energy. Here are some common abbreviations that university students use in academic lecture note-taking. Some abbreviations may have different meanings in various contexts, and there may also be variations in usage. Choose the ones you prefer; some people even create their own abbreviations.

Appendix B: Sample Notes of the Same Student (Pre-test at the top, Post-test at the bottom)

1. ^{at b}
 7. trad diet.
 6.
 7/11. snack
 mood

2. ^{of obvious}
 200 volun
 envi.
 original diet junk food

healthy diet ↔ healthy mind

eat diet ¹ west change
 ² oriental two times (b) -- (3)
 physical USA (for 20 times)

① 92 country
 1960-
 moderate
 Japan (3) remain?
 (b) low

② young men anger violent ph
 200 → 100 suppl. vitamin → # of violence
 100 previous not change
 after but before

healthy diet = healthy mind

Instructor Perceptions of Their Role in Student Motivation

Takako Moroi (Bunkyo Gakuin University)

Abstract

This study examines university instructors' perceptions of their role in student motivation and their strategies to foster engagement. As part of a larger research project, this qualitative study applies the MUSIC Model of Motivation and Theory X/Y climates to analyze how beliefs about responsibility shape teaching practices. Interviews with ten experienced instructors revealed three levels of perceived responsibility: strong, moderate, and realistic. Those with strong responsibility employed intensive, personalized strategies, while moderate responsibility instructors balanced collaboration and autonomy. Realistic responsibility instructors recognized external limitations while maintaining structured support. Findings highlight the importance of aligning professional development with instructors' perceived responsibility levels to enhance motivation and learning outcomes.

Keywords: instructor perceptions, motivational strategies, motivation climate, MUSIC model, student engagement

As educational landscapes continue to evolve, instructors have had to adapt their teaching methods, shifting from traditional in-person classrooms to emergency remote instruction and then back to physical classrooms. These transitions underscore the increasing importance of understanding motivational strategies across diverse learning environments (Bentley, 2020; Megahed & Ghoneim, 2022; Singh et al., 2021).

This study examines university instructors' perceptions of responsibility for student motivation and their motivational strategies. It is part of a larger research project investigating instructors' roles in student engagement across different educational contexts. By focusing on how instructors conceptualize their responsibility, this study contributes to the broader discussion on motivation in higher education,

highlighting key aspects of instructor motivation strategies that are consistent or adaptive across different instructional settings.

Conceptual Frameworks

This study is grounded in two key theoretical perspectives: the MUSIC Model of Motivation (Jones, 2009, 2018) and McGregor's (1960) Theory X and Theory Y. These perspectives provide a foundation for understanding instructors' perceptions of their responsibility for student motivation and the strategies they employ.

The MUSIC Model of Motivation and Motivation Climate

The MUSIC Model (Jones, 2009, 2018) identifies five essential components influencing student motivation: eMpowerment, Usefulness, Success, Interest, and Caring. Instructors who foster these elements create a learning environment that supports engagement. This study specifically focuses on how instructors' perceptions of responsibility influence their efforts to promote empowerment and success, as these components align with fostering student autonomy and self-efficacy.

Theory X and Theory Y Climates

McGregor's (1960) Theory X and Theory Y describe two contrasting views of motivation in the workplace, which can be applied to educational settings. Theory X assumes that individuals lack motivation and require external control, while Theory Y posits that people are naturally driven and thrive in environments that support autonomy. This study examines how instructors' beliefs about student motivation align with these theories, influencing whether they adopt directive (Theory X) or autonomy-supportive (Theory Y) teaching strategies.

By integrating these frameworks, this study explores how instructors conceptualize their role in student motivation and how these perceptions shape their instructional strategies.

Teachers' Perception of Student Motivation

The MUSIC Model of Motivation and Theory Y emphasize the role of instructors in fostering student motivation by creating supportive learning environments. The MUSIC Model outlines instructor actions that enhance motivation, while Theory Y frames student-centered teaching as essential for engagement. Research suggests that instructors influence student motivation through adaptive strategies

(Reeve & Cheon, 2016; Wang et al., 2017), making it crucial to understand factors that shape teachers' use of such strategies.

Laurermann and Karabenick (2013) developed the Teacher Responsibility Scale (TRS) to assess perceived responsibility in areas including student motivation. Studies applying the TRS consistently show that teachers rate their responsibility lowest for student motivation, often attributing it to internal student beliefs rather than their own instructional choices (Daniels et al., 2016, 2017, 2018; Eren, 2017). Teachers who view motivation as a stable characteristic may be less likely to implement motivational strategies, perceiving their impact as limited (Turner, 2010).

However, teachers' perceptions of student motivation are not always accurate. Ulstad et al. (2019) found that while teachers and students aligned in assessing autonomously motivated behaviors, teachers often misinterpreted passive behaviors as amotivation. Similarly, Schwan (2021) identified gaps in perception, with teachers attributing amotivation to external factors, while students cited a lack of relevance. Miller et al. (2017) observed that teachers with high self-efficacy rated remedial students as making greater progress in effort and achievement, suggesting that expectations may influence perceptions of motivation.

Teachers' beliefs about student motivation shape instructional choices (Berger et al., 2018; Liu et al., 2020). Hornstra et al. (2018) found that teachers with high expectations used more autonomy-supportive strategies, while Liu et al. (2020) demonstrated that teachers' views of motivation influenced their likelihood of using student-centered methods. These findings suggest that teacher expectations strongly impact motivational strategies and classroom environments.

Previous studies (Daniels et al., 2016, 2017, 2018) indicate that teachers often perceive student motivation as an internal trait rather than an instructional responsibility, which influences their engagement strategies (Liu et al., 2020). While research has explored this phenomenon in various educational contexts, little is known about how university instructors in Japan conceptualize their role in student motivation and how these beliefs shape their teaching practices. Addressing this gap, this study investigates the following research questions:

1. How do university instructors perceive their role in student motivation?
2. What strategies do university instructors use to foster student engagement, and how do these align with their perceived level of responsibility?

3. How do instructors assess the effectiveness of their motivational strategies, and what challenges do they encounter?

Method

This study was conducted as a qualitative phenomenological study to examine university instructors' perceptions of their role in student motivation. An interpretative phenomenological analysis (IPA) approach was chosen to provide an in-depth explanation, description, and exploration of the experiences and beliefs of instructors within a specific higher education context (Johnson & Christensen, 2020). This design allowed for the collection of rich, descriptive data to capture the complexity of how instructors view their responsibilities and influence student engagement.

Participants

The study included 10 participants with over 10 years of teaching experience to ensure they could provide detailed and reflective insights into their perceptions. They taught a range of skill-based English as a foreign language courses, including speaking, reading, writing, and standardized test preparation, at various higher education institutions in Japan. Participants were selected through purposeful sampling to examine their rich experience with a phenomenon of interest (Palinkas et al., 2015), meaning they were chosen specifically because their extensive teaching experience provided valuable insights into the phenomenon under study – university instructors' perceptions of responsibility for student motivation. Their perspectives were informed by years of active engagement in university-level teaching. Table 1 summarizes the participant demographics.

Table 1
Participant Demographics

Instructors	Teaching Experience	Gender	Employment Status	First Language
Instructor A	25 years	Female	Full-time	Japanese
Instructor B	22 years	Female	Full-time	Japanese
Instructor C	22 years	Female	Full-time	Japanese
Instructor D	36 years	Female	Full-time	English
Instructor E	20 years	Female	Full-time	English
Instructor F	12 years	Male	Full-time	English
Instructor G	21 years	Female	Full-time	Japanese/English
Instructor H	12 years	Male	Full-time	Japanese
Instructor I	10 years	Female	Full-time	Japanese
Instructor J	10 years	Female	Full-time	Japanese

Data Collection

Data were gathered using semi-structured interviews that aimed to uncover instructors' beliefs about their role in student motivation and the factors influencing these perceptions. The interview questions were designed to explore 1) the instructors' views on their responsibility in motivating students, 2) their perceptions of their influence on student engagement and learning outcomes, and 3) how these perceptions manifest in their teaching practices. In 2022, interviews were conducted via Zoom, a video conferencing tool. Each interview lasted approximately one to two hours and was audio-recorded with participant consent to ensure accuracy in transcription and analysis. The interviews were conducted in either Japanese or English, depending on the interviewee's preferred language. This approach ensured that participants could express their thoughts and experiences comfortably and accurately. The author then translated any quotes originally in Japanese into English for use in the study. This translation process was carefully undertaken to maintain the original meaning and nuances of the participants' responses.

Data Analysis

An IPA approach was used to interpret the data collected from the interviews. The analysis process began with an initial coding phase, identifying recurring concepts and ideas expressed by the participants (Larsen & Adu, 2021). Through multiple rounds of refinement, these codes were synthesized into core themes that captured instructors' perceptions and beliefs regarding their role in student motivation. Rather than applying a predetermined motivational framework, the analysis focused on understanding the depth and nuances of instructor perspectives, ensuring that their experiences were represented authentically.

In the coding process, key statements reflecting instructors' perceptions of motivational responsibility were identified. For example, one instructor stated, "I think the responsibility to motivate students lies with the instructor. I am confident that I can increase students' motivation, as it depends on what I do in class." This perspective was later recognized as aligning with a strong sense of responsibility. Another instructor explained, "I try different ways to engage students, but at the end of the day, their motivation has to come from within," reflecting a belief in shared responsibility. One instructor, in contrast, emphasized the influence of external constraints, stating, "I can provide structure and support, but some students simply won't engage, no matter what I do." These perspectives formed the basis for identifying broader patterns in instructors' views on motivation.

Results

This study examined university instructors' perceptions and strategies for motivating students through in-depth interviews. Analysis revealed three distinct levels of perceived responsibility: strong, moderate, and realistic. These categories reflect how instructors positioned their role in relation to external constraints and students' own motivation and responsibility for learning.

Instructors with a strong sense of responsibility saw themselves as primary drivers of student motivation. While acknowledging that students brought external influences, they emphasized the importance of their own instructional decisions during class time. Those with a moderate sense of responsibility viewed motivation as a shared effort, balancing structured support with student autonomy. They saw both instructor input and student initiative as essential. In contrast, realistic instructors emphasized the limits of their influence, placing greater weight on external factors that could affect engagement in unpredictable ways, even as they continued to offer support.

Table 2 presents a summary of the core beliefs and views of external factors associated with each group. While all instructors recognized that motivation is shaped by factors beyond their control, they varied in how much responsibility they assumed for addressing these challenges. The categories represent general tendencies rather than fixed identities, and in practice, motivational strategies sometimes overlapped. These categorizations reflect instructors' long-term teaching experience rather than temporary views shaped by a single course. Although some referenced specific classes, they did so to illustrate broader patterns, not course-bound approaches.

Table 2
Summary of Instructor Perceptions of Motivational Responsibility

Category	Core Belief	Views of External Factors
Strong	Instructors are primary drivers of motivation.	Acknowledged, but instructors focus on their own role within the classroom context.
Moderate	Motivation is shared between instructor and student.	Recognized as part of the dynamic.
Realistic	Instructors can support motivation, but influence is limited.	Emphasized and seen as shaping student engagement in unpredictable ways.

The following findings are organized according to these levels of perceived responsibility, illustrating how instructors positioned their role in student motivation, the strategies they employed, and the challenges they encountered. Table 3 provides a summary of instructors' perceived responsibility levels and the primary strategies they reported using.

Table 3
Instructor Perceptions of Responsibility for Student Motivation

Instructors	Perception of Responsibility Level	Perception of Responsibility	Strategies Used
Instructor A	Realistic	Acknowledges the limits of motivating all students	Direct teaching with motivational prompts; feedback cycles
Instructor B	Strong	Strong belief in active instructor influence	Task design that fosters participation; interactive activities
Instructor C	Moderate	Balanced; sees responsibility as shared between instructor and student	Engaging discussions; content tailored to interest
Instructor D	Strong	High commitment to fostering deeper connections and engagement	Storytelling; group-based learning activities
Instructor E	Strong	Focused on inclusivity and fostering broad engagement	Peer-led discussions; real-world applications
Instructor F	Strong	Focuses on creating high-energy, interactive activities to sustain student engagement and attention	Partner work; timed, active tasks
Instructor G	Moderate	Believes motivation is a shared effort, involving instructor and institutional support	Building rapport; collaborative projects
Instructor H	Strong	Believes in engaging through relatable, interesting content	Assignments that resonate with student interests; participation-focused activities
Instructor I	Strong	Believes motivation is a primary instructor responsibility	Personalized feedback, varied and adaptable teaching approaches

Instructor J	Moderate	Views motivation as supported by both instructor actions and institutional support	Clear communication of course goals; structured guidance
--------------	----------	--	--

Instructors with a strong sense of responsibility—Instructors B, D, E, F, H, and I—demonstrated a belief in their significant influence over student motivation. This belief was evident in their proactive teaching approaches and investment in maintaining high levels of student engagement. Instructor B described her role as crucial to creating an engaging classroom atmosphere: “I believe it’s my job to keep the class energy up. If I’m not engaged, how can I expect my students to be?” This mindset informed her task-oriented strategies that promoted active participation. Similarly, Instructor D leveraged storytelling as an engaging tool, emphasizing that her students responded positively, saying, “Students have told me they feel more connected to the material when I tell stories related to the lesson.” Instructor E incorporated inclusivity and peer-led discussions to create a sense of ownership among students, remarking, “Ensuring everyone feels involved is essential for sustained engagement.”

Instructors with a moderate sense of responsibility—Instructors C, G, and J—viewed their role as a partnership with students, emphasizing that motivation was a shared effort. Instructor C pointed out, “Motivation isn’t just on me. I try to create discussions that interest them, but they need to bring their part too.” This balanced approach led her to design discussions that prompted students to take an active role. Instructor G focused on building rapport to encourage student involvement, saying, “Building a connection with students helps, but they have to take some responsibility for staying engaged.” Instructor J saw her job as providing a structured environment with clear goals, noting, “My role is to set expectations, but students need to work within that framework to stay motivated.”

Instructor A, categorized under realistic responsibility, adopted a practical view that recognized the limits of her influence. She stated, “I’ve come to understand that even with the best prompts and feedback, some students won’t engage no matter how much I try.” She perceived her teaching strategies were effective for many students, but she acknowledged that factors outside her control could influence motivation. Among the instructors in this study, she was the only one who explicitly elaborated on the limitations of her influence on student motivation, making the realistic responsibility perspective unique to her. This perspective may stem from her extensive teaching experience of over 20 years, during which she

has encountered a range of student responses and learned to balance her motivational efforts with an understanding of external influences on student engagement.

These findings indicate that instructors' motivational strategies are closely tied to their perceived level of responsibility. Those with a strong sense of responsibility employ highly interactive and student-centered methods, while those with moderate responsibility emphasize collaboration. In contrast, instructors with a realistic perspective focus on structure and guidance while acknowledging external limitations. This distinction highlights the influence of instructors' beliefs on their teaching strategies and the role of responsibility perceptions in shaping classroom motivation.

Perceived Effectiveness and Challenges

Instructors employed varied strategies but faced challenges affecting their perceived effectiveness. Their assessments were based on observations of student engagement, including participation, feedback, and classroom behaviors. In this study, effectiveness refers to instructors' perceptions of how well their strategies motivated students and fostered engagement. Since student perspectives were not collected, effectiveness is entirely self-reported and reflects instructors' professional judgment.

To illustrate these perceptions, effectiveness levels were categorized based on instructors' assessments of student engagement. High perceived effectiveness describes strategies that consistently led to active participation and positive feedback, such as personalized feedback, interactive activities, and real-world applications. Moderate perceived effectiveness refers to methods that generally yielded positive results but required adaptation due to diverse student needs, external influences, or institutional constraints. Although limited perceived effectiveness was not observed in this study, future research may find this category useful for identifying strategies with minimal impact requiring refinement. Table 4 summarizes the participants' perceptions of their strategies' effectiveness and the associated challenges, offering a detailed view of their teaching practices.

Table 4

Instructors' Perceived Effectiveness and Challenges

Instructors	Perception of Responsibility Level	Perceived Effectiveness Level	Evidence Supporting Perceived Effectiveness	Challenges Affecting Engagement
--------------------	---	--------------------------------------	--	--

Instructor A	Realistic	Moderate	Positive feedback for most students; noted effectiveness in the majority	Acceptance that not all students will respond despite best efforts
Instructor B	Strong	High	Student comments like “time flies,” indicating high engagement	Initial low motivation in students; overcoming disengagement
Instructor C	Moderate	Moderate	Observable student engagement; gradual growth in intrinsic motivation	External factors impacting students’ motivation (e.g., home issues)
Instructors	Perception of Responsibility Level	Perceived Effectiveness Level	Evidence Supporting Perceived Effectiveness	Challenges Affecting Engagement
Instructor D	Strong	High	Positive feedback on storytelling and group work; strong student participation	Managing group dynamics and ensuring equal participation
Instructor E	Strong	High	Noted increased interaction and applied learning; students found lessons engaging	Balancing diverse student interests with relevant content
Instructor F	Strong	High	Rare student disengagement; high energy in class	Managing engagement in larger classes
Instructor G	Moderate	Moderate	Strong classroom rapport; active student participation	Dependence on institutional support, maintaining motivation outside class
Instructor H	Strong	High	High levels of involvement during tasks; responsive behavior	Adjusting content to meet diverse student interests
Instructor I	Strong	High	Positive student feedback; increased interest in English classes	Difficulty in reaching every student with a single approach

Instructor J	Moderate	Moderate	Noted increase in participation and interest among students	High expectations of institutional requirements affecting strategy implementation
--------------	----------	----------	---	---

The analysis in Table 4 provides insight into how the strategies used by instructors translated into perceived effectiveness and into the challenges they faced.

Instructors with high perceived effectiveness (B, D, E, F, H, and I) observed strong student engagement and received positive feedback. Instructor D emphasized storytelling: “Students have told me they feel more connected to the material when I tell stories.” Instructor B noted, “I often hear students say, ‘time flies’ during class, which tells me they’re involved.” Instructor E used real-world applications to enhance relevance: “Applying what we learn to real-world scenarios helps them see the value of their education.” Despite successes, challenges included managing large class participation (Instructor F) and adapting content for diverse student needs (Instructor H).

Instructors with moderate perceived effectiveness (A, C, G, and J) reported positive outcomes but faced limitations. Instructor A acknowledged, “Most of my students respond positively, but not everyone will engage.” Instructor C saw gradual improvements but cited external issues like home problems: “I’ve seen students improve, but external issues can set them back.” Instructor G emphasized rapport-building but noted, “Motivation sometimes drops without institutional support.” Instructor J highlighted balancing clear goals with institutional expectations: “Clear goals help, but meeting institutional demands can be difficult.”

No participants fell into the limited perceived effectiveness category, indicating that all instructors maintained at least moderate engagement levels. This finding underscores their adaptability and dedication.

The analysis reveals distinct patterns: high-effectiveness instructors engaged students through dynamic, personalized methods, overcoming challenges like class size and diversity. Moderate-effectiveness instructors achieved positive results but faced external barriers such as institutional constraints and external influences on motivation.

These findings illustrate how instructors’ perceived responsibility shapes their motivational strategies. Those who saw themselves as primary motivators employed proactive methods yielding high

engagement. Others, viewing motivation as shared or constrained by external factors, used effective strategies but faced unique challenges.

Perceived responsibility also influenced how instructors assessed their strategies. Those with strong responsibility found interactive approaches effective but noted engagement challenges. Moderate-responsibility instructors saw shared efforts as beneficial but faced variability in participation. Realistic-responsibility instructors emphasized structure and consistency while acknowledging external limitations.

The absence of limited perceived effectiveness reinforces the importance of aligning teaching strategies with beliefs about responsibility and fostering educational environments that support both instructors and students in achieving success.

Discussion

The results indicate that instructors' perceptions of their responsibilities shape their teaching strategies and their perceived effectiveness. Instructors with a strong sense of responsibility, such as Instructor D and Instructor B, used engaging, student-centered strategies like storytelling and interactive tasks, fostering high engagement. Instructors with a moderate sense of responsibility, such as Instructor C and Instructor G, balanced their strategies, recognizing motivation as a shared effort. The absence of participants in the limited effectiveness category highlights the dedication and adaptability of all instructors involved in this study.

Instructor's Role in Student Motivation

Student motivation is crucial in learning English and is shaped by teachers' practices and beliefs (Dornyei & Ushioda, 2021). Research has shown that teachers often report low levels of perceived responsibility for student motivation, particularly when they view it as innate and fixed (Daniels et al., 2016, 2017, 2018; Eren, 2017). However, Barahona and Darwin (2023) found that experienced teachers tend to feel more responsible than preservice teachers. In this study, many participants believed strongly in their motivational role, integrating strategies into daily practices. This finding aligns with Laurermann and Karabenick's (2013) concept of personal responsibility as an internal obligation and commitment beyond professional duty. The participants' extensive teaching experience likely contributed to their strong sense of responsibility, suggesting that such beliefs develop over time and through professional growth and time.

Challenges and Adaptations

The challenges noted in Table 2, such as diverse student needs and external factors, are consistent with the findings of Daniels et al. (2018), who reported that perceived external constraints can affect teachers' engagement in motivational practices. In this study, instructors with moderate perceived effectiveness, like Instructors A and C, highlighted those external factors, including students' home environments and institutional limitations, that influenced their ability to maintain motivation. This echoes Barahona and Darwin (2023), who emphasized that teachers' beliefs about responsibility are shaped not just by personal motivation but also by contextual factors and perceived support. Despite these challenges, the instructors demonstrated resilience by adapting their methods, as seen with Instructor D's storytelling approach and Instructor E's real-world applications, which were well-received by students. These findings underscore the importance of developing adaptive teaching practices and highlight the value of a supportive institutional framework that encourages innovative strategies.

Learning Environments

The findings of this study align with the principles of a positive motivation climate as outlined in the MUSIC Model (Jones, 2009) and reflect characteristics of the Theory Y learning environment (McGregor, 1960). Instructors with a strong or moderate sense of responsibility fostered autonomy, relevance, and connection – key elements of student engagement (Reeve & Cheon, 2016). This approach mirrors a Theory Y perspective, which emphasizes trust in students' intrinsic motivation and promotes agency and engagement (Hornstra et al., 2018). These instructors encouraged active participation through collaboration and responsiveness to student needs (Liu et al., 2020). The instructor with a realistic view acknowledged external constraints, stressing the need to balance motivational strategies with practical expectations. Overall, the findings support prior research on how instructor beliefs shape a student-centered learning environment (Daniels et al., 2018; Laurermann & Karabenick, 2013).

Practical Implications

Professional development programs that link intrinsic motivation with practical teaching strategies can strengthen instructors' sense of efficacy and commitment to fostering student engagement (Daniels et al., 2017; Eren, 2017). Aligning these programs with instructors' perceptions of responsibility can help institutions equip instructors to create motivational climates that support engagement across diverse educational contexts. The findings suggest that professional development should address the varying levels

of perceived responsibility among instructors. For those with a strong sense of responsibility, training that emphasizes innovative and student-centered pedagogical practices can enhance their already effective methods. Instructors with a moderate sense of responsibility may benefit from opportunities for collaboration with peers and supportive institutional policies that help them balance their shared view of responsibility with actionable strategies. Institutions can provide targeted resources and tools for instructors with realistic views to address external challenges while sustaining effective teaching practices. These tailored approaches can empower instructors at all levels to adapt their strategies and foster engaging, student-centered learning environments.

Limitations and Future Research

While this study offers valuable insights into instructors' perceptions and practices, its qualitative nature limits the generalizability of the findings. The analysis relied on participants' self-reported perceptions of the effectiveness of their teaching strategies, providing rich perspectives on instructor beliefs and classroom practices. However, the absence of direct student input or external validation of effectiveness constrains the breadth of these conclusions. Future research could employ quantitative approaches to measure teaching effectiveness and motivational strategies across larger, more diverse samples. Longitudinal studies examining how teachers develop their sense of responsibility for student motivation over time would further enhance understanding. Additionally, cross-cultural and comparative studies in different educational systems could explore the universality and contextual nuances of these motivational practices.

Conclusion

This study underscores the pivotal role of instructors in shaping student motivation through their perceptions of responsibility and the strategies they employ. By categorizing these perceptions into strong, moderate, and realistic views, the findings reveal how beliefs about teaching influence the creation of motivational climates in the classroom. This research highlights the importance of fostering autonomy, relevance, and connection in instructional practices, as supported by the MUSIC Model and Theory Y principles. Future research should explore how professional development initiatives can further support instructors in navigating the complexities of student motivation, ensuring positive learning outcomes in dynamic educational environments.

References

- Barahona, M., & Darwin, S. (2023). Balancing being a ‘good teacher’ and a ‘motivating teacher of English’: Analysing the sense of professional responsibility of pre-service EFL teachers. *Language Teaching Research*, 0(0). <https://doi.org/10.1177/13621688231183905>
- Bentley, K. (2020, April 15). Coronavirus: What are we learning about online instruction? *TCA Regional News*.
<https://tcsedsystem.idm.oclc.org/login?url=https://www.proquest.com/wire-feeds/coronavirus-what-are-we-learning-about-online/docview/2389737544/se-2?accountid=34120>
- Daniels, L. M., Poth, C., & Goegan, L. D. (2018). Enhancing our understanding of teachers’ personal responsibility for student motivation: A mixed methods study. *Frontiers in Education*, 3. <https://doi.org/10.3389/feduc.2018.00091>
- Daniels, L. M., Radil, A. I., and Goegan, L. D. (2017). Combinations of personal responsibility: Differences on pre-service and practicing teachers’ efficacy, engagement, classroom goal structures and wellbeing. *Frontiers in Psychology*, 8, 906.
<https://doi.org/10.3389/fpsyg.2017.00906>
- Daniels, L. M., Radil, A. I., and Wagner, A. K. (2016). Concordance between pre-service teachers’ responsibilities and instructional strategies. *The Journal of Experimental Education*, 84, 529–553. <https://doi.org/10.1080/00220973.2015.1054333>
- Dornyei, Z., & Ushioda, E. (2021). *Teaching and researching motivation* (3rd. ed.). Routledge.
- Eren, A. (2017). Investigating prospective teachers’ teaching-specific hopes as predictors of their sense of personal responsibility. *Asia-Pacific Journal of Teacher Education*, 45(3), 267–284.
<https://doi.org/10.1080/1359866X.2016.1169507>
- Hornstra, L., Stroet, K., van Eijden, E., Goudsblom, J., & Roskamp, C. (2018). Teacher expectation effects on need-supportive teaching, student motivation, and engagement: a self-determination perspective. *Educational Research and Evaluation*, 24(3-5), 324–345.
<https://doi.org/10.1080/13803611.2018.1550841>
- Johnson, R. B., & Christensen, L. (2020). *Educational research. Quantitative, qualitative, and mixed approaches* (7th ed.). SAGE

- Jones, B. D. (2009). Motivating students to engage in learning: The MUSIC model of academic motivation. *International Journal of Teaching and Learning in Higher Education*, 21(2), 272-285.
- Jones, B. D. (2018). *Motivating students by design: Practical strategies for professors* (2nd ed.). CreateSpace. <https://vtechworks.lib.vt.edu/handle/10919/102728>
- Larsen, H. G., & Adu, P. (2021). *The theoretical framework in phenomenological research: development and application*. Routledge. <https://doi.org/10.4324/9781003084259>
- Lauermann, F., & Karabenick, S. A. (2013). The meaning and measure of teachers' sense of responsibility for educational outcomes. *Teaching and Teacher Education*, 30, 13–26. <https://doi.org/10.1016/j.tate.2012.10.001>
- Liu, W. C., Wang, C. K. J., Reeve, J., Kee, Y. H., & Chian, L. K. (2020). What determines teachers' use of motivational strategies in the classrooms? A self-determination theory perspective. *Journal of Education*, 200(3), 185–195. <https://doi.org/10.1177/0022057419881171>
- McGregor, D. (1960). *The human side of enterprise*. McGraw Hill.
- Megahed, N., & Ghoneim, E. (2022). Blended learning: the new normal for post-COVID-19 Pedagogy. *International Journal of Mobile and Blended Learning*, 14(1), 1–15. <https://doi.org/10.4018/IJMBL.291980>
- Miller, A. D., Ramirez, E. M., & Murdock, T. B. (2017). The influence of teachers' self-efficacy on perceptions: Perceived teacher competence and respect and student effort and achievement. *Teaching and Teacher Education*, 64, 260–269. <https://doi.org/10.1016/j.tate.2017.02.008>
- Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2015). Purposeful Sampling for Qualitative Data Collection and Analysis in Mixed Method Implementation Research. *Administration and policy in mental health*, 42(5), 533–544. <https://doi.org/10.1007/s10488-013-0528-y>
- Reeve, J., & Cheon, S. H. (2016). Teachers become more autonomy supportive after they believe it is easy to do. *Psychology of Sport and Exercise*, 22, 178–189. <https://doi.org/10.1016/j.psychsport.2015.08.001>
- Schwan, A. (2021). Perceptions of student motivation and amotivation. *The Clearing*

House, 94(2), 76–82. <https://doi.org/10.1080/00098655.2020.1867490>

- Singh, J., Steele, K., & Singh, L. (2021). Combining the best of online and face-to-face learning: hybrid and blended learning approach for COVID-19, post vaccine, & post-pandemic world. *Journal of Educational Technology Systems*, 50(2), 140–171.
<https://doi.org/10.1177/00472395211047865>
- Turner, J. C. (2010). Unfinished business: putting motivation theory to the “classroom test.” In T. C. Urdan & S. A. Karabenick (Eds.), *Decade Ahead: Applications and contexts of motivation and achievement* (pp. 109–138). Emerald Group Publishing Limited.
[https://doi.org/10.1108/S0749-7423\(2010\)000016B007](https://doi.org/10.1108/S0749-7423(2010)000016B007)
- Ulstad, S. O., Halvari, H., & Deci, E. L. (2019). The role of students’ and teachers’ ratings of autonomous motivation in a self-determination theory model predicting participation in physical education. *Scandinavian Journal of Educational Research*, 63(7), 1086–1101.
<https://doi.org/10.1080/00313831.2018.1476917>
- Wang, C., Shim, S. S., & Wolters, C. A. (2017). Achievement goals, motivational self-talk, and academic engagement among Chinese students. *Asia Pacific Education Review*, 18(3), 295–307. <https://doi.org/10.1007/s12564-017-9495-4>

Structuring Polysemous Words' Instruction: A Hierarchical Approach

Hisanori Iijima (Graduate School, Keio University)

Key words: polysemous words, BERT, vocabulary acquisition, hierarchical approach

Abstract

Polysemous words, which possess multiple related meanings within a single word form, play a crucial role in second language acquisition (SLA) by enhancing learners' contextual meaning comprehension. However, previous research primarily examined high-frequency polysemous words, neglecting the semantic structures of low-frequency counterparts. This study investigated the clustering patterns of meanings in both high- and low-frequency polysemous words and proposed a pedagogically effective instructional framework. A set of polysemous words was randomly selected from Sugimori (2013)'s list based on frequency level and then classified into the JACET8000 and SVL12000 frequency bands. Sentences including these selected words were extracted from the English Wikipedia Corpus and analyzed using BERT-based contextual embeddings and hierarchical clustering based on cosine similarity. The results indicated that high-frequency polysemous words formed cohesive clusters with strong semantic connections, whereas low-frequency polysemous words exhibited more dispersed clusters with weaker meaning associations. These findings suggested that high-frequency polysemous words benefited from a domain-based instructional approach, while low-frequency polysemous words required a context-based methodology for effective comprehension. Based on these insights, this study introduced a frequency-based, hierarchical instructional framework, enabling educators to teach polysemous words by integrating semantic characteristics.

Introduction

Polysemy, the phenomenon in which a single word form possesses multiple related

meanings, is a fundamental feature of natural language. In second language acquisition (SLA), understanding polysemous words plays a crucial role in vocabulary knowledge, reading comprehension, and inferential abilities (Nation, 2023; Schmitt, 2020). Previous studies have primarily focused on high-frequency polysemous words, such as commonly occurring verbs, due to their frequent exposure in discourse and instructional materials (Meara, 2002; Schmitt & Schmitt, 2014). These words facilitate acquisition through repeated encounters, allowing learners to infer extended meanings from context (Nation, 2023).

In contrast, low-frequency polysemous words—lexical items that occur infrequently but exhibit multiple meanings—remain underexplored in SLA research (Schmitt, 2020). Words like *sanction* ("approval" vs. "punishment") and *strain* ("variety" vs. "tension") exemplify this issue, as their meanings exhibit weaker semantic relatedness, making them harder to acquire (Chronis & Erk, 2020). Given their rarity, incidental learning is less likely, necessitating explicit instruction (Webb, 2008). Despite their low occurrence, these words are essential in academic and professional contexts, highlighting the need for pedagogical strategies tailored to their learning.

Although traditional linguistic approaches, which rely on dictionary definitions and manually annotated corpora, struggle to capture the context-dependent nature of low-frequency polysemy, for the recent advancements in natural language processing (NLP), particularly contextual embedding models like BERT (Devlin et al., 2019), can offer a data-driven alternative by analyzing large-scale linguistic patterns. These models reveal how polysemous meanings cluster within language use, providing insights into structural differences between high- and low-frequency words.

This study investigates how polysemous word meanings are clustered across different frequency levels using BERT-based contextual embedding analysis. By examining these structural differences, it aims to inform vocabulary instruction strategies that integrate frequency-based considerations. By bridging computational methodologies and pedagogy, this

research offers empirical insights into polysemy acquisition and instruction.

Literature Review

Defining High- and Low-Frequency Polysemous Words

Polysemy, a widely studied topic in linguistics and SLA, refers to a single word possessing multiple related meanings. The lexical frequency hypothesis (Ellis, 2002) suggests that high-frequency words are acquired and retained more easily due to repeated exposure. However, how high- and low-frequency polysemous words differ in acquisition remains unclear.

High-frequency polysemous words such as *run* and *take* appear across diverse contexts, which reinforce strong semantic associations through distributed learning (Webb, 2020; Nation, 2023). Vocabulary instruction strategies often leverage this by clustering related meanings to aid retention (Laufer & Nation, 1995; Schmitt & Schmitt, 2014).

However, low-frequency polysemous words present distinct challenges. Their meanings exhibit weaker semantic cohesion, making contextual inference unreliable (Chronis & Erk, 2020). Limited exposure and weak conceptual connections further hinder learners' ability to generalize meanings (Wei & Lou, 2015). Consequently, explicit instruction becomes crucial.

NLP-Based Approaches to Polysemy

Recent NLP advances provide systematic methods for polysemy analysis. BERT (Devlin et al., 2019), a powerful contextual embedding model, captures semantic variations based on linguistic context. Empirical studies indicate that high-frequency polysemous words form tight semantic clusters, whereas low-frequency words exhibit more dispersed representations, which reflects weaker conceptual coherence (Chronis & Erk, 2020). These findings align with cognitive linguistic perspectives, suggesting that high-frequency words rely on core meanings, while low-frequency words require inferencing.

Despite BERT's advantages over traditional models like Word2Vec (Mikolov et al.,

2013) and GloVe (Pennington et al., 2014), it often struggles with low-frequency words due to insufficient training data. It also fails to distinguish metaphorical extensions from literal meanings (Periti & Tahmasebi, 2024), indicating the need for improvements such as diachronic modeling and domain-specific corpora.

Evaluating Polysemous Vocabulary Knowledge: BERT-Assisted Assessments

Assessing polysemy comprehension remains a challenge. Traditional vocabulary tests tend to prioritize breadth over depth, making them inadequate for evaluating polysemy. BERT-based assessments, leveraging cosine similarity measurements, offer a scalable solution by analyzing learners' word usage in context (Liu et al., 2022). These AI-driven tools enhance the precision of polysemy evaluation, offering new avenues for SLA pedagogy.

Research Gaps and Objective

Despite extensive research, low-frequency polysemous words remain underexamined in SLA, particularly regarding cognitive processing and pedagogy. While NLP models have advanced meaning disambiguation, their integration into SLA instruction is still in its early stages.

This study addresses these gaps by analyzing meaning clustering patterns in high- and low-frequency polysemous words using BERT-based contextual embeddings. Additionally, it evaluates the effectiveness of explicit instruction for low-frequency polysemy and proposes an AI-enhanced instructional framework. This framework integrates NLP-based assessment tools with evidence-based SLA pedagogy, aiming to optimize vocabulary instruction.

Research Objective / Questions

This study investigates how word frequency influences the semantic clustering of polysemous words using BERT-based contextual embedding analysis and cosine similarity metrics. Additionally, it develops frequency-sensitive teaching strategies based on empirical finding. The research questions are as follows:

1. How do the meaning clustering patterns of low-frequency polysemous words, as measured through cosine similarity-based clustering in contextual embedding analysis, differ from those of high-frequency polysemous words?
2. What pedagogical implications can be drawn from frequency-based patterns of polysemous word structures?

Methodology

This study analyzes the semantic clustering of polysemous words with the use of English Wikipedia corpus. Wikipedia corpus was chosen due to its broad topical coverage, balanced mix of naturalistic language use and controlled vocabulary, and accessibility, making it an appropriate resource for examining the distribution of polysemous word senses across various domains.

To ensure statistical validity, 1,000 sentences per target word were stratified and randomly sampled to achieve a balanced representation of different senses, minimizing bias in meaning distribution. This approach ensures that each word's contextual usage is proportionally represented in the dataset, providing a robust foundation for semantic analysis.

Polysemous words were selected from Sugimori (2013)'s list, referencing frequency data from JACET8000 (2003) and further validated using Mizumoto (2022)'s New Word Level Checker. To ensure contemporary relevance, these words were compared against SVL12000 (ALC, 2010) which reflect modern language usage patterns. Analysis showed that 78.2% of Sugimori's listed words remained within the top 3,000 in recent frequency lists. Prior research suggests that vocabulary within this range constitutes the core learning set for language learners (Nation, 2001). Based on this threshold, the study classified words into high-frequency ($\leq 3,000$) and low-frequency ($> 3,000$) categories. From these categories, five high-frequency and twenty low-frequency words were randomly sampled for analysis. This selection ensured a diverse representation of polysemous words across frequency bands, facilitating a

balanced examination of their semantic clustering patterns.

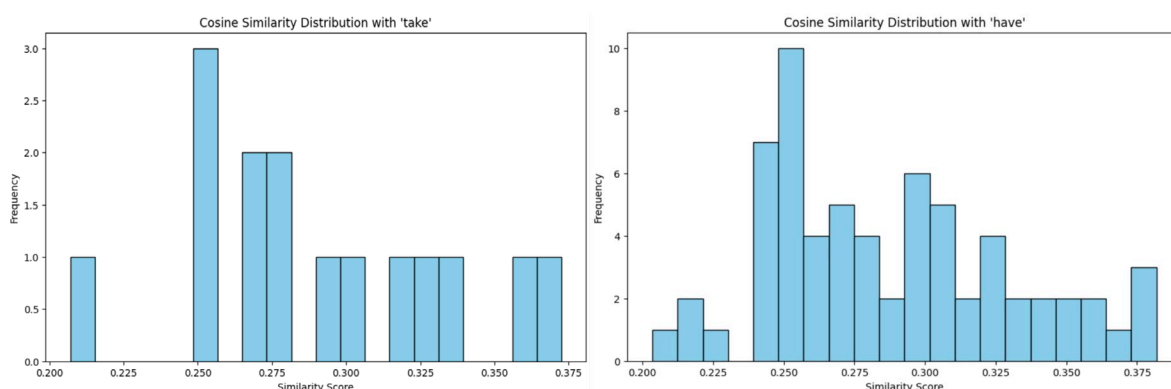
To maintain a balanced dataset, each word was randomly selected to avoid selection bias. To verify the polysemy of selected words, WordNet was used to ensure that each had at least two distinct senses, following Sugimori (2013). This verification step confirmed that the dataset accurately represented words with multiple meanings. The alignment between corpus-based contextual meanings and WordNet’s sense distinctions was assessed using BERT-based embeddings, with cosine similarity (threshold: 0.75) employed to evaluate semantic distinctiveness, ensuring that identified meanings were meaningfully distinct rather than ambiguous variations.

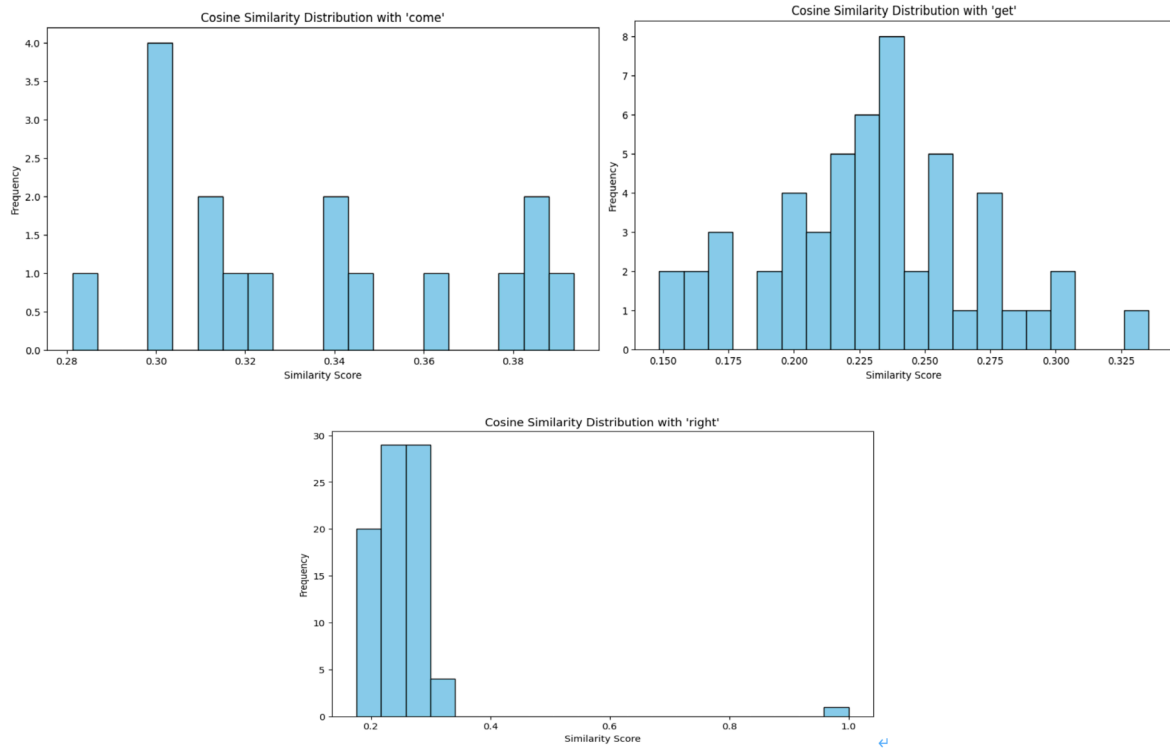
For semantic clustering analysis, BERT-large (uncased) was employed, fine-tuned using Masked Language Modeling (MLM) and Next Sentence Prediction (NSP) to enhance its contextual understanding.

Results

High-frequency Polysemous Words Analysis

The analysis of cosine similarity distributions for five high-frequency polysemous words—"have," "take," "get," "right," and "come"—revealed distinct distributional patterns in their semantic representations.





(Figure 1: Cosine Similarity Distribution with high-frequency polysemous words)

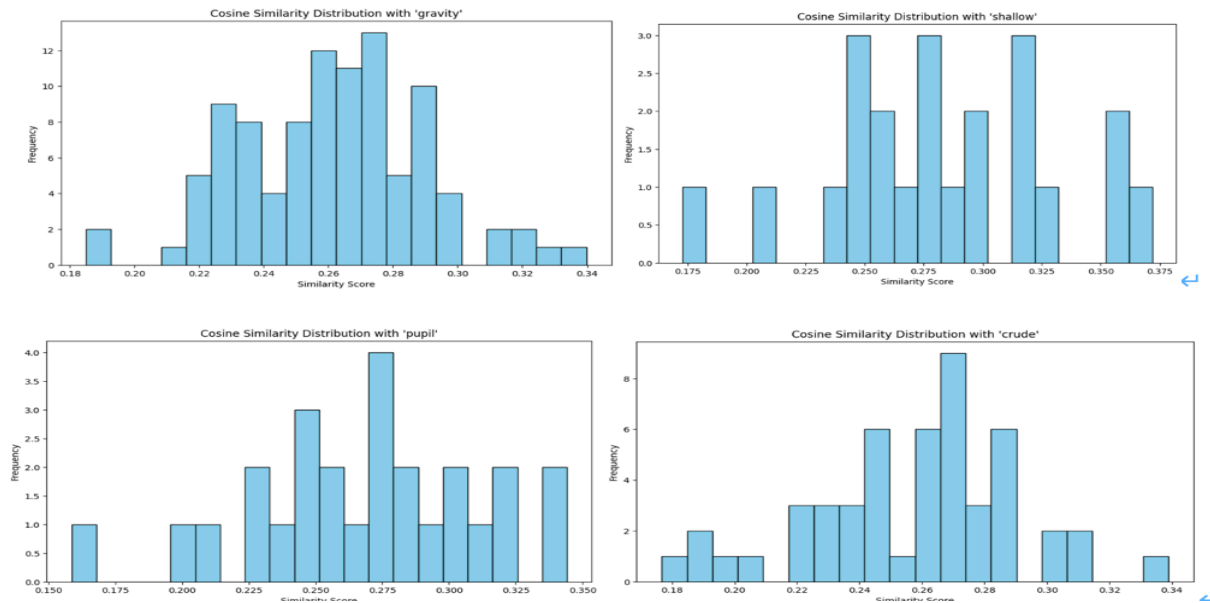
The verb "have" exhibited a prominent peak at approximately 0.25 in its cosine similarity distribution, with 68% of its usage instances clustering within the 0.20-0.30 range. Similarly, "take" showed a concentrated distribution pattern with a peak at 0.25, where 72% of instances fell within the same range. The word "get" displayed a broader distribution spanning from 0.15 to 0.325, with multiple smaller peaks observed across this range, forming three distinct clusters: 0.15-0.20 (31%), 0.20-0.25 (42%), and 0.25-0.325 (27%). The word "right" demonstrated a highly concentrated distribution around 0.2, with 85% of instances falling within the 0.15-0.25 range. In contrast, "come" exhibited multiple peaks within the 0.30-0.38 range, with the distribution split across three main clusters: 0.30-0.32 (35%), 0.32-0.35 (40%), and 0.35-0.38 (25%).

Low-frequency Polysemous Words Analysis

The analysis of low-frequency polysemous words revealed four distinct distribution patterns. Among the twenty words, missing values were identified in 4 cases, and these were

excluded from the analysis.

The first pattern, the physical-metaphorical dichotomy, was observed in words such as "shallow," "gravity," "pupil," and "crude". which exhibited bimodal distributions.

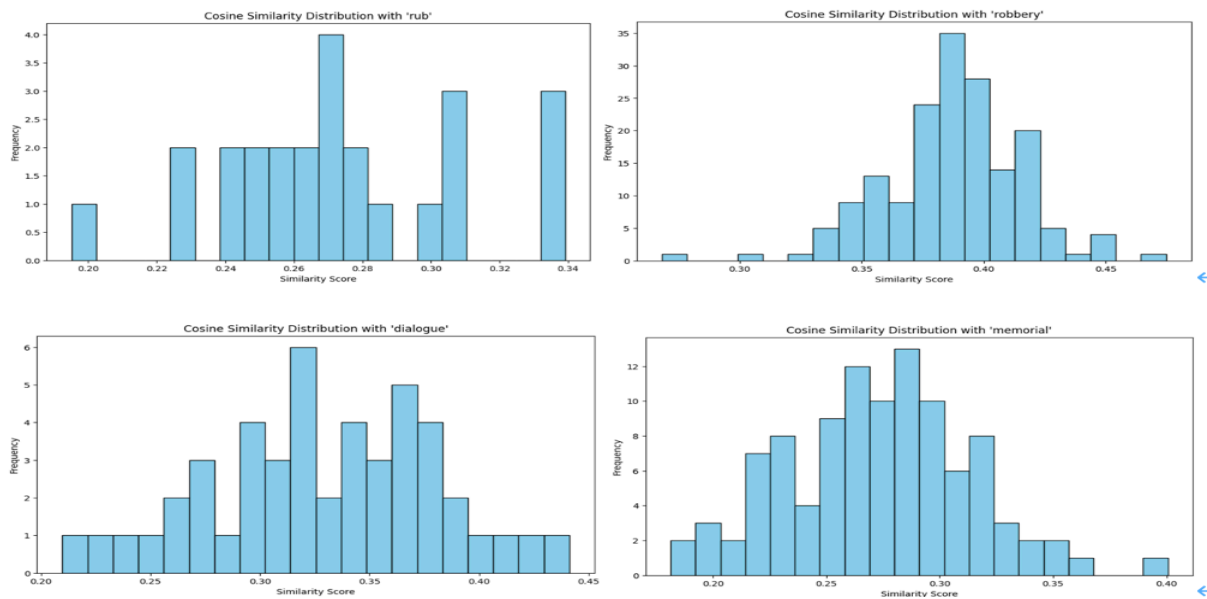


(Figure 2: Cosine Similarity Distribution with “shallow,” “gravity,” “pupil,” and “crude”)

The first peak appeared in the 0.15-0.25 range, representing physical meanings commonly associated with tangible objects or directly perceivable phenomena. These meanings tend to be more concrete, reflecting sensory experiences such as depth, weight, or spatial properties. In contrast, the second peak, occurring in the 0.35-0.45 range, corresponds to metaphorical meanings that extend beyond physical properties into abstract or figurative domains. These metaphorical meanings often emerge through conceptual mappings, where a physical concept is used to structure an abstract idea. The clear separation between these peaks suggests that the two categories are cognitively distinct, with minimal blending between them, reinforcing the hypothesis that metaphorical meanings develop independently from their physical counterparts.

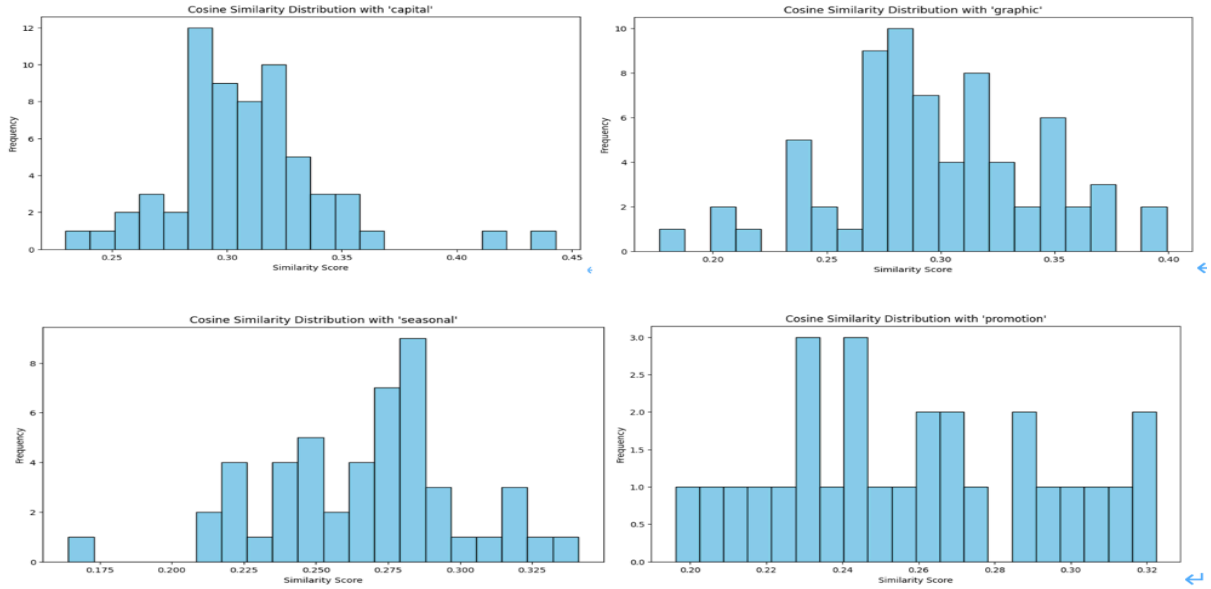
The second pattern, context-constrained distribution, was evident in words such as "rub," "robbery," "memorial," and "dialogue", which demonstrated a narrow distribution range of

0.20-0.30 with a single prominent peak.



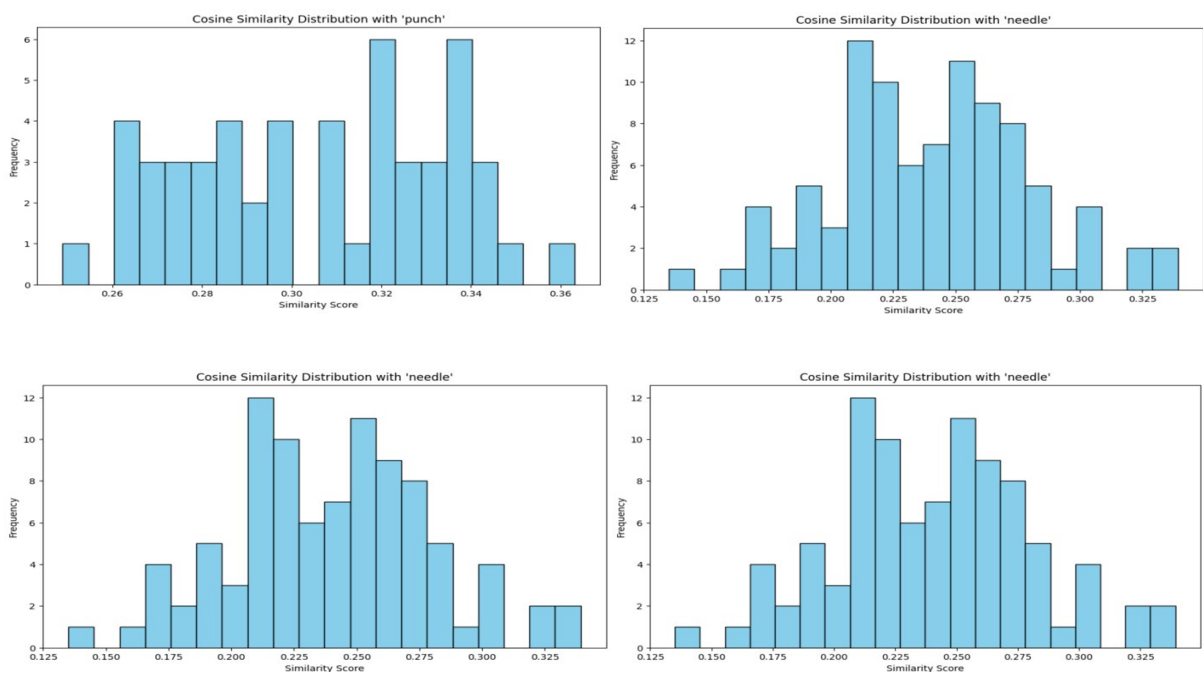
(Figure 3: Cosine Similarity Distribution with "rub," "robbery," "memorial," and "dialogue")

The high concentration of 80-90% of instances within a 0.05 range indicates that these words possess a well-defined and constrained semantic scope. This suggests that their meanings are not highly variable or context-dependent but rather remain stable across different usages. Such words tend to appear in specialized or domain-specific contexts where their meaning is clearly understood and rarely shifts beyond a narrow range. In contrast, the third pattern of gradual semantic expansion, as seen in words like "seasonal," "promotion," "capital," and "graphic," reveals a more flexible and evolving semantic structure. The continuous distribution across the 0.15-0.40 range suggests that these words undergo meaning extension in an incremental manner, adapting to various contextual nuances over time. Rather than having sharply defined peaks, the smooth transitions in their distribution imply that different meanings blend into one another, allowing for a gradual conceptual shift that accommodates new interpretations while retaining connections to their core meanings.



(Figure 4: Cosine Similarity Distribution with "seasonal," "promotion," "capital," and "graphic")

The fourth pattern, disjoint polysemy, was found in words such as "needle," "riot," "punch," and "journal," which exhibited multiple distinct peaks with clear separations between them.



(Figure 5: Cosine Similarity Distribution with "needle," "riot," "punch," and "journal")

These words exhibited distinct semantic clusters that remained largely separate, indicating that their multiple meanings are not closely related or easily interchangeable. This lack of overlap suggests that the meanings are contextually dependent and do not form a cohesive semantic network. The wide distribution range, spanning from 0.10 to 0.50, further reinforces the idea that these words encompass highly varied meanings, some of which may stem from historical changes, domain-specific usage, or even accidental polysemy. Such a pattern aligns with cases where words acquire multiple meanings due to linguistic evolution rather than systematic semantic extension. As a result, learners may struggle to infer connections between these meanings, necessitating explicit instruction to distinguish them and clarify their appropriate contextual applications.

Discussion

Overall Discussion

This study investigated the impact of word frequency on the semantic clustering of polysemous words, revealing distinct differences in their structural and distributional patterns. The findings indicate that high-frequency polysemous words form cohesive and well-defined clusters, whereas low-frequency counterparts display fragmented and contextually variable distributions. These results provide valuable insights into polysemy in second language acquisition and support the development of frequency-sensitive instructional strategies. The discussion addresses two key research questions: (1) How do the clustering patterns of low-frequency polysemous words, as measured by cosine similarity-based contextual embedding analysis, differ from those of high-frequency words? and (2) What pedagogical implications arise from these frequency-based semantic structures?

RQ1: Differences in Meaning Clustering Patterns Between High- and Low-Frequency

Polysemous Words

The results indicate that high-frequency polysemous words exhibit more cohesive semantic structures, characterized by narrowly concentrated cosine similarity distributions and a strong central peak that serves as a semantic anchor. This is particularly evident in words such as *have* and *take*, where 68–72% of usage instances are clustered within a narrow similarity range (cosine similarity: 0.20–0.30). Such findings suggest that high-frequency words maintain stable and interconnected semantic representations, reinforcing the hypothesis that frequent exposure facilitates the consolidation of core meanings and their systematic extensions. Moreover, words such as *get* exhibit a broader yet systematically structured distribution, with multiple peaks within a relatively small range. This pattern suggests a process of systematic meaning extension, where distinct yet related senses emerge through recurrent linguistic use. These observations are consistent with usage-based theories of language acquisition (Ellis, 2002), which argue that frequent usage reinforces the stability and interconnectedness of semantic networks.

In contrast, low-frequency polysemous words demonstrate considerably more diverse and fragmented distributional patterns, reflecting the varied semantic structures that arise due to differences in their cognitive and contextual representations. The analysis revealed four primary distributional patterns among low-frequency words. First, words exhibiting a physical-metaphorical dichotomy, such as *shallow* ("not deep" vs. "superficial") and *gravity* ("gravitational force" vs. "seriousness"), show bimodal distributions with clearly separated peaks, suggesting minimal overlap between their physical and metaphorical meanings. This pattern aligns with Conceptual Metaphor Theory (Lakoff & Johnson, 1980), which posits that metaphorical meanings develop independently from their concrete counterparts.

Second, context-constrained words, such as *robbery* and *memorial*, demonstrate highly concentrated distributions, with 80–90% of their occurrences falling within a narrow similarity range (cosine similarity: 0.20–0.30). These words are predominantly used in domain-specific

contexts, which restricts their semantic variation. Interestingly, this finding challenges Zipf's (1935) law of meaning distribution, which predicts greater semantic variability in lower-frequency words. Instead, the results suggest that some low-frequency words maintain stable meanings due to their specialized and contextually restricted usage.

Third, words following a pattern of gradual semantic expansion, such as *promotion* ("advancement" vs. "marketing strategy") and *capital* ("seat of government" vs. "financial asset"), exhibit continuous distributions without distinct peaks (cosine similarity: 0.15–0.40). This suggests a progressive broadening of meanings over time, in line with the Semantic Network Model (Taylor, 2003), which theorizes that meaning extension occurs incrementally through exposure to diverse linguistic contexts.

Finally, disjoint polysemous words, such as *needle* ("sewing tool" vs. "to provoke") and *punch* ("a strike" vs. "a beverage"), exhibit multiple non-overlapping peaks (cosine similarity: 0.10–0.50), indicating historically unrelated meanings. This pattern corresponds with Pustejovsky's (1995) theory of accidental polysemy, which suggests that certain words acquire multiple meanings through historical and external linguistic influences, rather than through systematic semantic development.

Overall, these findings provide strong empirical support for the hypothesis that high-frequency words develop cohesive, interconnected semantic structures, while low-frequency words display more fragmented and diverse meaning distributions. This distinction underscores the necessity of frequency-sensitive approaches to vocabulary instruction, as discussed in the following section.

RQ2: Pedagogical Implications of Frequency-Based Semantic Structures

The structural differences between high- and low-frequency polysemous words necessitate distinct instructional strategies to enhance vocabulary acquisition. Given their strong core meanings and systematic extensions, high-frequency polysemous words benefit from network-based instruction, which emphasizes the structured introduction of extended

meanings through semantic mapping. By utilizing high-frequency contexts from corpora, learners can reinforce core meanings while gradually internalizing extended senses through exposure to authentic linguistic materials. Additionally, pattern recognition activities that encourage learners to identify systematic meaning variations can further enhance their ability to infer novel senses.

In contrast, low-frequency polysemous words require a more tailored, pattern-specific instructional approach that accommodates their diverse semantic structures. Words exhibiting a physical-metaphorical dichotomy necessitate explicit instruction that highlights the relationship between their concrete and abstract meanings. The use of visual aids and metaphor awareness activities can facilitate comprehension by making conceptual mappings more salient. For context-constrained words, domain-specific instruction that incorporates specialized texts and professional discourse is essential for clarifying their restricted applications. Words that follow gradual semantic expansion should be introduced sequentially, with semantic mapping employed to illustrate the progressive development of meanings. In the case of disjoint polysemy, where meanings are historically unrelated, each sense should be treated as an independent lexical entry, with explicit contextual differentiation to prevent confusion. Emphasizing usage-based distinctions in authentic contexts enables learners to develop a more precise and contextually appropriate understanding of these words.

Beyond its pedagogical implications, this study contributes to the broader field of second language vocabulary acquisition by providing empirical evidence of frequency-based semantic clustering patterns. The findings challenge traditional dictionary-based representations of polysemy, highlighting the necessity of corpus-based and computational approaches to polysemy research. However, several limitations must be acknowledged. The reliance on BERT for semantic clustering analysis, while valuable, does not fully replicate human-like semantic interpretation, necessitating further research that integrates behavioral data to validate cognitive processing mechanisms. Additionally, as this study focuses

exclusively on English, its findings may not be directly applicable to morphologically rich languages such as Japanese or Finnish, where polysemy may operate differently. Future research should examine cross-linguistic variations in frequency-based polysemy structures. Finally, although this study proposes a frequency-sensitive instructional framework, its effectiveness remains untested in pedagogical settings. Future studies should incorporate experimental validation in classroom environments to assess the practical impact of these instructional strategies.

Limitation

This study faces three main limitations. First, while BERT models offer insights into semantic relationships, they cannot fully replicate human semantic interpretation, particularly with low-frequency words and cultural implications. Second, the study's focus on English and its limited sample size (five high-frequency and twenty low-frequency words) from Wikipedia may not adequately represent polysemy across languages and contexts. Third, the proposed instructional framework, though theoretically grounded, requires empirical validation through classroom-based studies to assess its effectiveness across different proficiency levels and first language backgrounds.

Conclusion

This study examined how word frequency influences the semantic clustering of polysemous words, revealing that high-frequency words form cohesive clusters, while low-frequency words exhibit dispersed and contextually distinct meanings. These findings highlight the need for frequency-sensitive instructional strategies tailored to each word type.

For high-frequency polysemous words, integrative teaching methods such as semantic mapping and meaning extension are effective, as their meanings are closely connected. In contrast, low-frequency polysemous words require explicit instruction, including contrastive

analysis and domain-specific categorization, to clarify their diverse meanings. Using authentic materials, such as news articles and academic texts, enhances learners' ability to interpret these words in context.

The proposed hierarchical instructional framework provides a structured approach to polysemous vocabulary learning by aligning teaching methods with word frequency. While this study focuses on English, the principles may apply to other languages. Future research should validate these strategies in classroom settings and explore the role of computational models, such as NLP-based semantic clustering, in improving polysemy instruction.

In conclusion, adopting a frequency-based instructional approach enhances vocabulary acquisition by providing targeted, contextually relevant learning experiences. By implementing structured teaching methods, educators can help learners achieve deeper comprehension, better retention, and more effective usage of polysemous words.

Notes

1. The experiments utilizing BERT were conducted using a program developed and verified on the latest version of Python.
2. The 20 target words analyzed in this study as low-frequency polysemous words are as follows: "snap", "dialogue", "quotation", "slot", "capital", "crude", "shallow", "journal", "punch", "rub", "gravity", "memorial", "riot", "seasonal", "tense", "pupil", "needle", "promotion", "robbery", "graphic".
3. The high polysemous words analyzed in this study are “take”, “get”, “have”, “right”, “come”.

Acknowledgements

I would like to express my deepest gratitude to the anonymous reviewers for their invaluable contributions to this paper, as well as their thoughtful and constructive feedback, which has greatly helped improve the quality of this work and provided me with new insights

and perspectives for future research. Their dedication and expertise have been truly inspiring.

References

- ALC. (2010). *SVL12000: Standard Vocabulary List 12,000*. Retrieved November 5, 2024, from <http://www.alc.co.jp/vocogram/article/svl/>
- Boers, F., & Lindstromberg, S. (Eds.). (2008). *Cognitive linguistic approaches to teaching vocabulary and phraseology*. Mouton de Gruyter.
- Chronis, G., & Erk, K. (2020). When is a bishop not like a rook? When it's like a rabbi! Multi-prototype BERT embeddings for estimating semantic relationships. *Proceedings of the 24th Conference on Computational Natural Language Learning*, 227–244.
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT)*, 4171–4186.
- Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, 24(2), 143–188.
- JACET. (2003). *JACET8000: JACET list of 8000 basic words*. Japan Association of College English Teachers.
- Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. University of Chicago Press.
- Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, 16(3), 307–322.
- Liu, Q., Sun, H., Liu, Y., Qiu, X., & Huang, X. (2022). Semantics-aware BERT for word sense disambiguation. *Artificial Intelligence*, 301, 103586.
- Meara, P. (2002). The rediscovery of vocabulary. *Second Language Research*, 18(4), 393-407.
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed

- representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems (NeurIPS)*, 26, 3111–3119.
- Mizumoto, A. (2022). *New Word Level Checker*. Kansai University. Retrieved November 5, 2024, from <https://nwlc.pythonanywhere.com/>
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge University Press.
- Nation, I. S. P. (2023). *How vocabulary is learned (2nd ed.)*. Oxford University Press.
- Pennington, J., Socher, R., & Manning, C. (2014). GloVe: Global vectors for word representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543.
- Periti, F., & Tahmasebi, N. (2024). A systematic comparison of contextualized word embeddings for lexical semantic change. *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, 4262–4282. Association for Computational Linguistics.
- Pustejovsky, J. (1995). *The generative lexicon*. MIT Press.
- Schmitt, N. (2020). *Vocabulary in language teaching (2nd ed.)*. Cambridge University Press.
- Schmitt, N., & Schmitt, D. (2014). A reassessment of frequency and vocabulary knowledge. *Language Learning*, 64(3), 487–514.
- Sugimori, N. (2013). WordNet wo mochiita eigo tagigo risuto no kochiku [Construction of an English polysemous word list using WordNet]. *Ritsumeikan Language and Culture Studies*, 24(4), 171–182.
- Taylor, J. R. (2003). *Linguistic categorization: Prototypes in linguistic theory (3rd ed.)*. Oxford University Press.
- Webb, S. (2008). The effects of context on incidental vocabulary learning. *Reading in a Foreign Language*, 20(2), 232–245.
- Webb, S. (2020). *Researching vocabulary: A vocabulary research manual*. Palgrave Macmillan.
- Wei, X., & Lou, Y. (2015). A study of word meaning acquisition of polysemous words for English

majors. *Creative Education*, 6(18), 1993–2000.

Zipf, G. K. (1935). *The psycho-biology of language: An introduction to dynamic philology*.

Houghton Mifflin.

How ChatGPT Improves Writing for EFL Students

Sumie Matsuno(Aichi Sangyo University College)

Key words: ChatGPT, English as a foreign language, paragraph writing, university, Comparison and contrast

Abstract

The paper discusses the use of ChatGPT in an English writing class at a Japanese university, focusing on a comparison and contrast writing task. Thirty-four students, with above-average English proficiency, were asked to compare their original writing with ChatGPT-revised versions. The results showed that students appreciated ChatGPT's improvements in vocabulary, grammar, and sentence structure, with many noting clearer and more polished writing. However, there were concerns about ChatGPT misunderstanding students' original intentions, often making unwanted changes to content and emotional tone. Some students highlighted issues like the omission of personal experiences or specific examples, while others observed improved organization and flow in their paragraphs. Overall, the study revealed that ChatGPT is a useful tool for enhancing English writing, but students must critically evaluate its revisions to ensure their intended message is preserved. The research concludes that educators should guide students in using AI effectively while fostering their independent language skills. This initial experiment with ChatGPT was deemed successful, though the teacher plans to explore further applications of AI in the classroom to maximize its benefits.

Introduction

The rise of AI tools like ChatGPT has posed challenges for teaching English writing in EFL

(English as a Foreign Language) classes. Students frequently rely on ChatGPT to complete their writing assignments, often without putting in the necessary effort. Universities are increasingly advising teachers to assign tasks that cannot be easily completed using AI or to conduct in-class tests instead. However, in a writing-focused classroom, where activities like drafting, revising, and proofreading are essential, in-class tests may not be suitable. In this paper, I will present a teaching approach used in my writing class that addresses these challenges.

Literature Review

The reviewed studies in 2024 examine the integration of ChatGPT in education, emphasizing both its benefits and challenges. Yalcin and Serkan (2024) found that graduate students often accepted ChatGPT's information without verification, highlighting the need for educators to guide students on using AI critically rather than avoiding it. Firth, Derendinger, and Triche (2024) developed a framework for appropriately using AI tools like ChatGPT, tested in MBA and undergraduate business classes. It aims to balance AI benefits with preventing misuse, fostering engaging classroom discussions. Gurhan and Serkan, C (2024) surveyed academics on their views of ChatGPT. Despite mixed emotions like anxiety and hopefulness, most were open to using the tool in the future, emphasizing the need for guidelines in its use. Selahattin & Eyup (2024) integrated ChatGPT into the Flipped Learning model, enhancing personalized learning and active student participation, especially in math and science subjects. Studies by Kyle (2024) in Korea and Phuong (2024) in Vietnam revealed that ChatGPT improved student engagement and language skills, though some preferred traditional instruction and faced technical challenges. Ekrem (2024) and Tseng and Lin (2024) explored ChatGPT's role in language education, highlighting its potential for individualized feedback and skill enhancement. However, challenges like maintaining academic integrity and ensuring effective integration into teaching remain. Li (2024) reviewed literature on ChatGPT's role in second language acquisition, emphasizing the importance of improving AI literacy among both students and teachers to maximize its benefits.

Overall, these studies collectively underscore the need for a thoughtful, balanced approach to AI integration in education. While AI tools like ChatGPT offer numerous benefits, their successful application hinges on fostering critical thinking, providing teacher training, and addressing the ethical implications surrounding their use. Given the concerns about using AI, this study aims to explore how its practical application can affect students' writing skills.

Participants

Thirty-four university students attended this study. Their major was literature, and this university is one of the top-level national universities in Japan. Hence, their English level was above that of the ordinal Japanese university students. I assume that they had a B1 or B2 level in the CEFR framework (Common European Framework of Reference for Languages).

Procedures

The students were assigned to write a comparison and contrast paragraph about their high school and university life. After completing their paragraphs, they brought them to class for group discussions, where each group member's paragraph was reviewed. Following the group discussion, they were asked to input their paragraphs into ChatGPT. Then, they were tasked with writing a comparison and contrast paragraph between their original paragraphs and the ChatGPT-revised versions. The following outlines the instructions I gave my students.

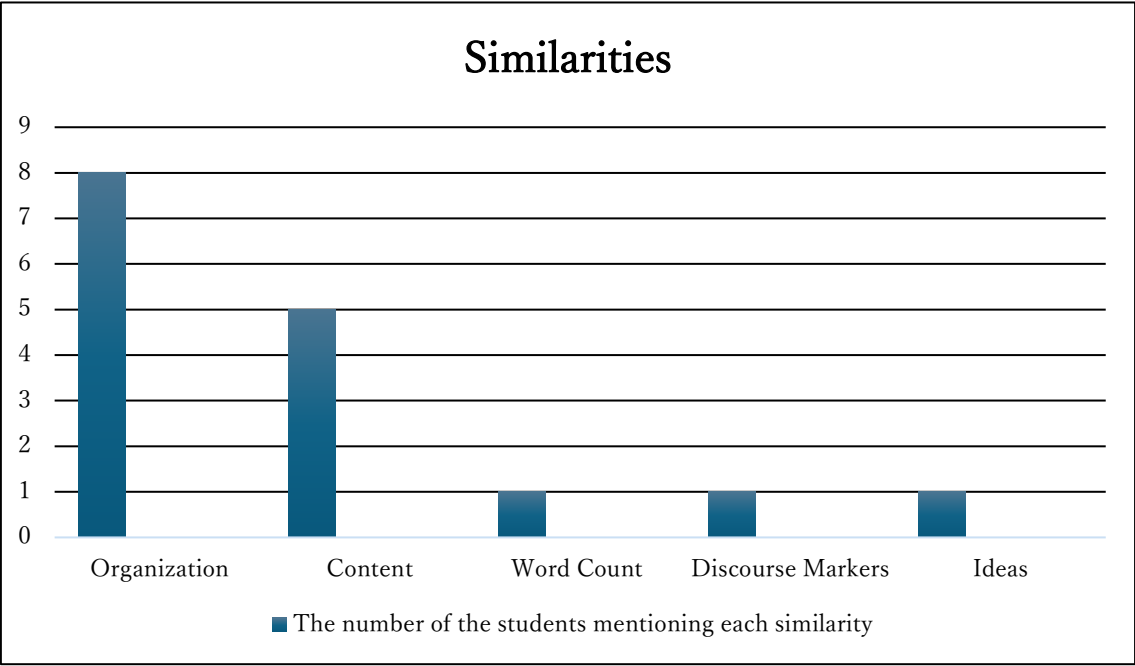
Access ChatGPT and enter the following command: "Please revise the following paragraph." After that, copy and paste your original paragraph into the ChatGPT. Then, compare your original paragraph with the paragraph revised by ChatGPT and write a comparison and contrast paragraph. Also, include the contents of your peer review in the paragraph (250 words or more).

At the end of the class, students were asked to submit their original paragraph, ChatGPT's revised version, and their final comparison paragraph. Unfortunately, two students misunderstood the instructions and did not submit their final paragraph. As a result, a total of 32 students submitted their final work. In this paper, I will analyze their final paragraphs and examine how they compared their original and ChatGPT-revised versions.

Results and Discussions

When reviewing the students' final comparison and contrast paragraphs between their original work and the one revised by ChatGPT, I noticed that all students explained the differences, but only 16 students highlighted the similarities. Figure 1 presents the similarities between the original paragraphs and ChatGPT's versions.

Figure 1
Similarities between the original paragraph and ChatGPT's version.



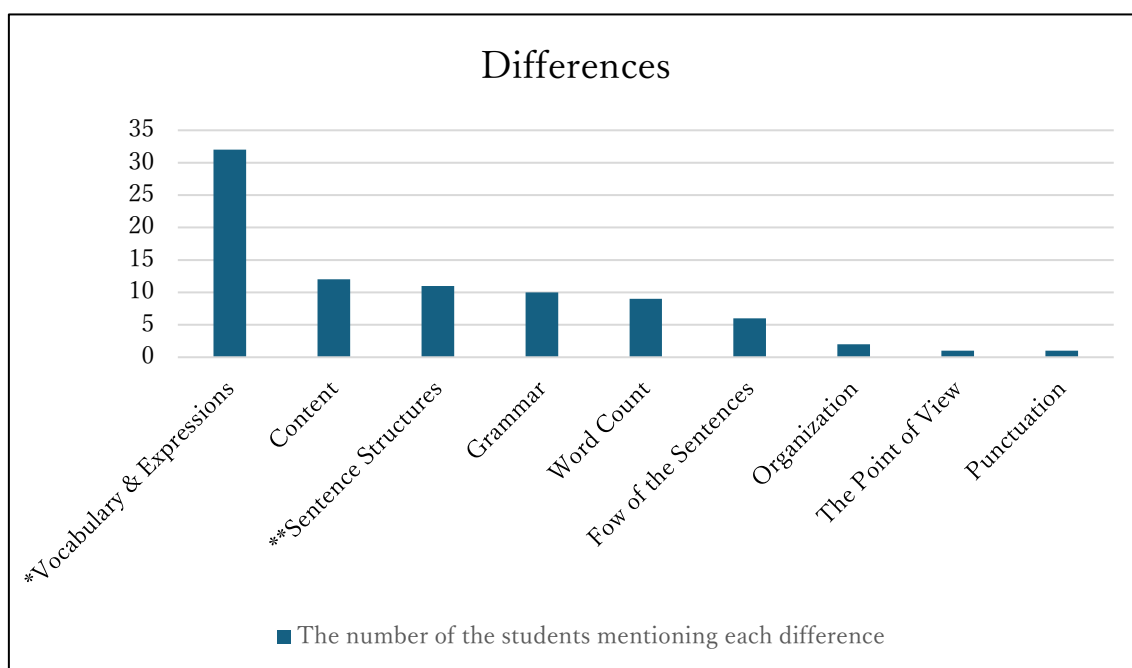
Eight students mentioned that the organization of their original paragraph was not revised. Five

students noted that the content remained the same. One student observed that the word count was nearly identical, stating, "According to my count, the original paragraph had 200 words, while the revised one had 199. I found this interesting because the two paragraphs used different expressions." Another student pointed out that ChatGPT did not alter her discourse markers, while one more mentioned that the ideas in both paragraphs were identical.

Figure 2 shows the results of the differences. As you can see, all the students (32 out of 32 students) noted that ChatGPT changed the vocabulary and expressions in their paragraphs. ChatGPT used a wide variety of vocabulary and avoided repeating words, whereas the students tended to repeat the same words, which ChatGPT replaced. In fact, 10 students specifically mentioned the issue of repeated words in their original paragraphs. Additionally, eight students mentioned that proper conjunctions, such as "for instance," were added to their original paragraphs, making the writing clearer and easier to read. One student also pointed out that "ChatGPT did not use hedges like 'may' or 'might,' which indicates that AI is more objective and tends to eliminate subjective expressions."

Figure 2

The differences between the original paragraph and ChatGPT's version.



*Vocabulary & Expressions include discourse markers.

**Sentence Structures include sentence length.

Regarding the content, 12 students mentioned that ChatGPT made some changes to their original paragraphs. Three students highlighted positive aspects of these changes:

"ChatGPT uses specific examples, like classes, clubs, and study abroad opportunities, which helps make my opinion clearer."

"ChatGPT adds more details, such as explaining the responsibilities and skills needed in college."

"ChatGPT removes redundant information."

It appears that ChatGPT provides general details to enhance clarity and eliminates unnecessary information it deems irrelevant.

On the other hand, nine students highlighted negative aspects of using ChatGPT, primarily due to misunderstandings by the AI. One student shared that she had written extensively

about her personal experiences, but ChatGPT shortened that section. Another student noted that ChatGPT misunderstood his intent, resulting in a revision that changed the meaning of his sentences. Similarly, one student observed that ChatGPT adopted an objective tone, minimizing her personal experiences by removing emotional elements.

Another expressed frustration that ChatGPT replaced specific examples with more general information. One student reported that the AI misinterpreted his writing, leading to inaccuracies. Another remarked, "The way emotions are expressed is different. AI is just a tool that imitates human intellectual abilities." One student pointed out that ChatGPT removed sentences with concrete numbers, which impacted the clarity of their comparison. Another student wanted to compare high school and university classes separately in terms of length and quantity, but ChatGPT confused the two, incorrectly stating that college classes are twice as long as high school classes. One student was disappointed that ChatGPT misinterpreted her intent, revising her paragraph to focus on hobbies and traveling instead of balancing studying with those activities. Lastly, one student commented, "The original paragraph becomes more correct and simpler. However, what I really wanted to say may have been lost in the pursuit of brevity."

Considering these concerns, I assume that ChatGPT may have omitted personal experiences or specific examples because it tends to default to a neutral or formal tone, particularly in academic or professional writing. As a result, it may remove emotional or subjective elements, making personal experiences appear less relevant. Additionally, this omission might stem from ChatGPT's inherent limitations, as it cannot fully grasp or replicate the writer's personal identity and lived experiences.

Eleven students commented on sentence structure. Some noted that their original sentences were short and simple, while ChatGPT produced more complex sentences using conjunctions, relative pronouns, and advanced grammar. One student observed, "The length of the sentences increased after revision. AI's paragraph now includes several relative pronouns and participial constructions, resulting in a smoother overall flow." Conversely, another student stated,

“ChatGPT’s sentences are shorter and clearer than mine,” while another remarked, “The sentences were summarized and much easier to read.” One student noted that ChatGPT utilized various sentence structures that he hadn’t employed. Another student indicated that the length of ChatGPT’s sentences is long, but there are fewer sentences overall.

Grammar was another area where ChatGPT often made revisions. In fact, ten students mentioned that their grammatical errors were corrected by ChatGPT. One student stated, “I struggle with using the present tense, past tense, and present perfect tense appropriately, so ChatGPT revised them.” Another noted, “I made many mistakes in my use of prepositions.”

In terms of word count, nine students noted that ChatGPT produced shorter paragraphs than their originals. This was primarily due to ChatGPT’s removal of redundant information and unnecessary expressions, resulting in condensed paragraphs. A comparison of word counts between the original and ChatGPT-revised paragraphs revealed that ChatGPT consistently shortened the students’ writing. One student stated, “The word count of my original paragraph is 353, while that of the revised paragraph is 273.” Another student noted, “The length of the paragraph decreased from 246 words to 193 words.”

Regarding the flow of sentences, one student remarked that her original paragraph had a somewhat complicated flow, whereas ChatGPT’s version was more orderly. Another student mentioned that the revised version strengthened the logical connections between sentences, making transitions smoother and the overall argument more coherent. One student observed, “Each sentence logically follows the previous one, enhancing the flow of the narrative.” Another commented, “The ChatGPT paragraph maintains a clearer and more logical structure.”

Only two students noted that their organization was revised by ChatGPT. Both mentioned that the paragraph structure became clearer and easier to follow. Regarding the point of view, one student remarked that ChatGPT changed it, stating, “The point of view became more objective.” As for punctuation, one student wrote, “Punctuation was also addressed, with the revised text using commas and periods more effectively.”

Their comments about CHATGPT

Not all students commented, but 18 provided feedback about ChatGPT. The following are their comments:

- "I was able to discover various similarities and differences."
- "ChatGPT is better than I thought."
- "I doubt whether AI can create perfectly correct sentences."
- "ChatGPT can write more orderly and easier-to-understand paragraphs than I do."
- "ChatGPT improved my paragraph in some ways but also made it worse. This shows that while ChatGPT can be useful, it can also pose a threat depending on how we use it."
- "The ability to revise text in ChatGPT is impressive, so I think we should make good use of it."
- "I found that ChatGPT's paragraph is much easier to read compared to mine."
- "By using ChatGPT, we can create more polished paragraphs. However, there are risks involved, so it's important to use it effectively to enhance our academic skills."
- "ChatGPT provided me with techniques to improve my sentences."
- "Overall, it seems that AI produces better paragraphs."
- "I was impressed by ChatGPT's convenience and believe we should use it wisely."
- "ChatGPT helped me identify my weaknesses in writing English paragraphs."
- "The paragraphs created by ChatGPT feature more advanced vocabulary and sentence structures; however, my intended message can be difficult to convey."
- "Thanks to the comparison, I made many discoveries."
- "Using ChatGPT to write reports may not be ideal, but I would like to use it to check my sentences."
- "AI is probably superior to me in writing, and the development of AI is remarkable; I

have much to learn from it."

- "I was surprised that ChatGPT can create more natural English sentences than many non-native speakers."
- "I've decided to use ChatGPT for writing English essays."

Based on the comments, students generally found ChatGPT to be a helpful tool for improving their writing, particularly in terms of organization, vocabulary, and sentence structure. Many students noted that ChatGPT produced clearer, more polished, and easier-to-read paragraphs than their own. However, they also recognized its limitations, especially in accurately conveying their intended message. While some students appreciated the convenience and advanced capabilities of AI, they emphasized the importance of using it wisely to complement their own skills rather than relying on it completely. Overall, they saw ChatGPT as a valuable resource for improving their writing, but also acknowledged the need for careful review and critical thinking when using it.

Conclusion

I found this writing assignment quite successful because the students recognized noticeable improvements in their writing. Additionally, some students realized that ChatGPT sometimes misunderstood their intended meaning. As a result, after receiving ChatGPT's revised paragraphs, they had to carefully check whether the changes accurately conveyed their original intentions.

Furthermore, this study highlighted that while ChatGPT effectively changes vocabulary and expressions, it isn't flawless when it comes to content. Therefore, I need to remind my students to be mindful of this aspect.

Overall, this generation has access to AI, which is vastly different from my own school days. Since we cannot ignore AI, it's crucial for teachers to guide students in using it effectively, while also helping them improve their own language skills. This was my first attempt at

incorporating ChatGPT into my class, and although it was quite successful, I still need to explore additional ways to make the most of AI.

References

- Ekrem, S. (2024). Revolutionizing Language Learning: How ChatGPT and AI Are Changing the Way We Learn Languages. *International Journal of Technology in Education*, 7(2), 353-372.
- Firth, D.R., Derendinger, M., & Triche, J. (2024). Cheating Better with ChatGPT: A Framework for Teaching Students When to Use ChatGPT and Other Generative AI Bots. *Information Systems Education Journal*, 22(3):47-60.
- Gurhan, D., & Serkan, C. (2024) Are Alarm Bells Ringing in Academia? ChatGPT as a Sample of Using Chatbots in Education. *Turkish Online Journal of Distance Education*, 25(3), 1-17.
- Kyle, R.V.H (2024). ChatGPT in English Language Learning: Exploring Perceptions and Promoting Autonomy in a University EFL Context. *TESL-EJ*, 28(1).
- Li, Y. (2004). Capabilities, effectiveness, applications, challenges, and solutions. *Studies in Applied Linguistics & TESOL*, 24(1), 24-37.
- Phuong, H.P.X. (2024). Using ChatGPT in English language learning: A Study on I.T. students' attitudes, habits, and perceptions. *Online Submission*, 4(1):55-68.
- Selahattin, A., & Eyup, Y. (2024) Flipped Learning: An Innovative Model for Enhancing Education through ChatGPT. *International Journal of Modern Education Studies*, 8(1), 124-148.
- Tseng, Y., & Lin, Y. (2024). Enhancing English as a Foreign Language (EFL) Learners' Writing with ChatGPT: A University-Level Course Design. *Electronic Journal of E-Learning*, 22(2), 78-97.
- Yalcin, D., & Serkan, B. (2024) From 'Can AI think?' to 'Can AI help thinking deeper?': Is Use of ChatGPT in higher education a tool of transformation or fraud? *International Journal of Modern Education Studies*, 8(1), 49-71.

**教職課程の学生の英会話能力向上を目指した大学英語授業における
多様な学習活動とその効果**

**The effect of varied activities for the development of EFL speaking
performance among students of a teacher certification program**

瀧澤典子（宇都宮大学）

キーワード：教員養成，大学英語授業，英会話能力，英語学習活動，自己効力感

Abstract

This study examines the effect of tailored activities aimed at improving EFL speaking performance among student teachers enrolled in an English-language communication course as part of a teacher certification program at a Japanese university ($N=29$). A pre-survey revealed that over two-thirds of the students considered speaking to be the skill they felt least proficient among the four core English language skills. To address this, from April to July 2022, the course incorporated practical conversation activities, such as quizzes and a mini debate as well as chants that help to improve speaking fluency. These activities, some of which were adapted from a communication skills coursebook, were introduced in three phases to support the students' gradual development of speaking skills. To evaluate the effect of this pedagogical practice, both pre- and post-surveys were conducted, alongside speaking tests for four individual students. The evaluation summary from four English language instructors indicated that while average speaking self-efficacy remained low, the aforementioned four students either maintained or improved in speech intelligibility, especially in pronunciation and lexical coherence. Notably, the post-survey indicated that students found learning English response expressions particularly effective in developing their speaking communication skills.

1. はじめに

現行の学習指導要領の「(付録 8)『外国語活動・外国語の目標』の学校段階別一覧表」(文部科学省, 2018) には、小学校、中学校、高等学校の外国語活動および外国語について、目標言語でコミュニケーションを図る資質・能力を「言語活動」を通して育成することが目標とされている。ここで言われる「言語活動」とは言語材料について理解し機械的な練習することとは区別され、「実際に英語を使用して互いの考えや気持ちを伝え合う」(文部科学省, 2017, p.23) ことを指し、各学校段階において外国語活動および外国語の授業を担う教員もまた、児童生徒と円滑かつ柔軟に意思疎通をとる英語力、つまり英語コミュニケーション能力が求められる(文部科学省, 2019)。小学校英語が全面実施されて 5 年が経とうとしているが、現職の小学校教員のみならず教員養成課程の学生も、依然として外国語指導に関するさまざまな不安を払拭できないでいることが先行研究において指摘されている(津田, 2022)。

小学校英語全面実施前の 2015 年、小学校教員 174 名を対象に行なわれた小学校外国語活動の教科化・低学年化に対する不安に関する意識調査(米崎, 多良, 佃, 2016) の分析結果から、教科化について、「教員の英語力・指導力」、「国語や他教科とのバランス」、「児童の負担・混乱」に関する不安が抽出された。「教員の英語力・指導力」については英語発音力への不安、専門性への不安、「国語や他教科とのバランス」については母語習得への不安、「児童の負担・混乱」については苦手意識を持つことへの不安、学習負担の増大への不安が含まれる。また、これらに加えて評価内容の不明瞭さや、評価による子どもたちへの負の影響から、「評価への不安」も顕在化された。

そして、小学校英語の全面実施後、現職の小学校教員 134 名を対象に行われた「小学校の英語教育に関する教員意識調査 2021」の結果(イーオン, 2021) によると、外国語を担当したことのある教員($N=57$)のうち、指導についてうまくいっていない、またはあまりうまくいっていないと感じている教員は 42%を占め、その半数以上が「スピーキング」、とりわけ話すこと[やりとり]の指導に困難を抱えていることが分かった。さらに、同年に実施された千葉県教員($N=90$)に対する英語指導上の不安に関する調査(田

中, 2023) からも、評価方法に次ぎ、音声指導、自身の発音に不安を感じている教員が 50%以上を占めたことが報告されている。教員の発音に関する不安は全面実施前の調査 (米崎他, 2016) においても抽出されているが、「発音に敏感な時期に間違った発音で指導すれば子どもたちの発音は間違ったものになる」(p.139) ことに加え、英語学習の初期段階にある児童が「読む」、「書く」技能を習得していく上でも、教員には音素の正確な発音と文字を対応させた指導が求められる。

また、2021 年から 2022 年にかけて教員養成課程の学生 121 名を対象に実施された小学校英語に対する不安に関する意識調査の結果 (津田, 2022) によると、英語の授業を行うことに対する前向きな期待と同時に、授業を行うことに対する不安を抱いている学生が 80%以上であることが分かった。その不安の内容は「英語に対する苦手意識と自信のなさから生じる不安」、「話す際の発音と文法への不安」、「英語の知識のなさへの不安」、「ALT の先生とのコミュニケーションについての不安」の 5 つに分類されているが、津田は「根本的に英語に対する苦手意識が指導に対して多大な不安を引き起こしている」(p.125) こと、さらに、必ずしもこの外国語不安 (苦手意識) が実際の英語力と合致していない可能性も指摘している。

以上の先行研究から、小学校の外国語活動および外国語が全面実施されて以降も現職の小学校教員のみならず教員養成課程の学生の意識には、英語の授業を行うことに対する不安、とりわけ言語活動を通して児童の英語コミュニケーションの資質・能力を育成する上で必要となる英会話能力に関する不安が根強く残っており、その克服が求められる。そこで、本研究では小学校教員養成課程の学生を対象に必修科目である「コミュニケーション英語」の授業において、英会話能力に関する不安の克服とその向上を目的とした実践研究を行うこととした。

2. 研究の概要

2.1 研究の目的

本研究の目的は、小学校教員養成課程の必修科目「コミュニケーション英語」の授業において学習者の英会話能力の向上を目的とした多様な学習活動を継続的に行い、その

教育的効果と課題を探ることである。これに従い、次の3つの研究課題を設定した。第1の研究課題は、当該授業における学習活動を通して受講者がどの程度発音能力を向上させることができたのかを明らかにすることである。第2の研究課題は、当該授業における学習活動を通して受講者がどの程度英会話能力を向上させることができたのかを明らかにすることである。第3の研究課題は、当該授業全体を通して受講者がどのような学習活動を英会話能力の向上に役立つと感じたかを明らかにすることである。

2.2 対象者

本研究の対象者は日本の大学における小学校教員養成課程の必修科目である「コミュニケーション英語」受講者の大学生29名である。そのうち応募法により、英語教員免許取得を目指す学生3名および国語教員免許取得を目指す学生1名を抽出者とした。

2.3 調査方法

当該科目における学習者の小学校英語教育に対する自己効力感およびニーズを把握すること、また、その変化をはかることを目的とし、本研究は以下に説明する手順で調査を実施した。まず、非英語母語話者の受講者（ $N=38$ ）に対し任意のアンケートによる事前事後意識調査を行った。次に、前述の第1、第2、第3の研究課題に関する調査の一環として、応募法による抽出者4名に対し事前事後の英語スピーキングテストおよびインタビューを実施した。そして、スピーキングテストについては、学習者の英語運用能力の客観的評価データを得ることを目的とし、第三者の立場にある英語教員4名に評価を依頼した。各調査方法の詳細は表1の通りであり、事前事後のアンケートによる意識調査の質問項目については、現行の『小学校学習指導要領（外国語活動・外国語編）』（文部科学省, 2019b）に記載されている、聞くこと、読むこと、話すこと[やりとり]、話すこと[発表]の4技能5領域に関する目標および内容のうち、話すこと[やりとり]、話すこと[発表]および英語の音声に関する事項を反映させた（表2）。これにより、小学校英語教育に求められるスピーキング・コミュニケーションに関する能力について、項

目ごとに対象者の自己効力感をはかった。なお、対象者の同意は、調査目的と概要とともに記載されているアンケート上の確認項目の回答により得た。

表 1：研究調査内容の詳細

事前意識調査	
実施日	2022 年 4 月 13 日
対象者	受講者 38 名中 29 名
内容	Google フォームを用いて、「発音」について 4 項目、「話すこと（やりとり）」について 3 項目、「話すこと（発表）」について 3 項目、「書くこと」について 2 項目の質問から成る 5 段階評定尺度形式（1 全くそう思わない、2 そう思わない、3 どちらでもない、4 そう思う、5 非常にそう思う）のアンケートを実施
事前テスト	
実施日	2022 年 4 月 25 日
対象者	受講者 38 名中、応募による抽出者 4 名（20 歳男性 2 名、20 歳女性 2 名）
内容	非英語母語話者の日本人英語教員 1 名が、各対象者に対し、短い物語文（ <i>The Boys and the Frog</i> ）に関する 3 つのタスク（Part A 物語文の音読、Part B 即興を前提とするやりとり、および Part C ある程度の準備をした上で話すこと）から成る、対面による 10 分間の英語スピーキングテストを実施
事後意識調査	
実施日	2022 年 7 月 11 日
対象者	受講者 38 名中 28 名
内容	Google フォームを用いて、事前意識調査の全項目に加え、「当該科目の授業内容」について 2 項目の質問から成る 5 段階評定尺度形式（1 全くそう思わない、2 そう思わない、3 どちらでもない、4 そう思う、5 非常にそう思う）および補足説明を求める記述式の 1 項目を含むアンケートを実施
事後テスト	
実施日	2022 年 7 月 19 日、7 月 20 日、7 月 21 日
対象者	事前テストと同参加者 4 名
内容	事前テストと同様（ <i>The Ant and the Dove</i> ）
事後インタビュー	
実施日	2022 年 7 月 19 日、7 月 20 日、7 月 21 日
参加者	事前テストと同参加者 4 名
内容	非英語母語話者の日本人英語教員 1 名が、各対象者に対し、スピーキング・コミュニケーション能力の向上のための授業内外における学習活動について 5 項目の質問から成る、対面による 10 分間の半構造化インタビューを実施

事前事後の対面による英語スピーキングテストについては、抽出者 4 名の実用英語技能検定（英検）取得級に鑑みた内容とし、構成については平成 26 年度に文部科学省が実施した英語力調査（文部科学省、2016）を参考にした。事前テストではイソップ物語（Jones, 1912）から *The Boys and the Frogs* を、事後テストでは同物語から *The Ant and the Dove* を題材とし、第 1（Part A）に音読、第 2（Part B）に即興的なやりとり、そして第 3（Part C）に準備した上での発表（意見）を行うことを課題とした。そして、続く事後インタビューについては、各抽出者に対して当該科目の授業を通して英会話能力が向上したかどうか、授業においてどのような学習活動がその向上に役立つと感じたか、また、授業外に取り組んでいる英語学習について問う半構造化インタビューを実施した。なお、英語スピーキングテストおよびインタビューは当該科目の授業担当者が所属する研究機関の

倫理審査委員会の審査を経るとともに（承認番号: 2022-71）、対象者の承諾を得た上で IC レコーダーを用いた録音を行った。

表 2：事前事後意識調査回答の統計分析結果

項目番号	項目内容	事前 (N=29)		事後 (N=28)		<i>p</i> 値	効果量 <i>d</i>
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
I 発音							
I-1	現代の標準的な発音ができる	3.07	1.11	3.21	1.15	.67 ns	.12 小
I-2	語と語の連結による音の変化をつけることができる	3.14	1.20	3.32	1.10	.58 ns	.16 小
I-3	語や句、文における基本的な強勢を使い分けられる	3.14	1.04	3.11	1.05	.11 ns	.03 小
I-4	文における基本的なイントネーションを使い分けられる	3.21	1.06	3.25	.99	.14 ns	.04 小
II 話す（発表）							
II-1	日常生活に関する身近で簡単な事柄について、簡単な語句や基本的な表現を用いて話すことができる	3.66	.92	3.79	.86	.54 ns	.15 小
II-2	自分のことについて、伝えようとする内容を整理した上で、簡単な語句や基本的な表現を用いて話すことができる	3.62	.96	3.86	.79	.49 ns	.28 小
II-3	身近で簡単な事柄について、伝えようとする内容を整理した上で、自分の考えや気持ちなどを、簡単な語句や基本的な表現を用いて話すことができる	3.45	1.00	3.68	.89	.49 ns	.24 小
III 話す（やりとり）							
III-1	基本的な表現を用いて指示、依頼をしたり、それらに応じたりすることができる	3.48	1.00	3.75	.83	.31 ns	.29 小
III-2	日常生活に関する身近で簡単な事柄について、自分の考えや気持ちなどを、簡単な語句や基本的な表現を用いて伝え合うことができる	3.79	.89	3.82	.85	.98 ns	.03 小
III-3	自分や相手のこと及び身の回りの物に関する事柄について、簡単な語句や基本的な表現を用いてその場で質問をしたり質問に答えたりして、伝え合うことができる	3.34	.99	3.75	.87	.11 ns	.44 中

† $p < .10$ * $p < .05$ ** $p < .01$

$d = 0.2$ （効果量小）、 $d = 0.5$ （効果量中）、 $d = 0.8$ （効果量大）

2.4 授業実践

初回授業で実施した事前意識調査の結果を受け、学習者のニーズに基づき英会話能力向上を支援すべくさまざまな学習活動を全 15 回のコースに段階的に導入した（表 3）。

まず、イントネーションや音の連結、強勢等、発音に関する自己効力感（表 2: 項目 I-1; 項目 I-2; 項目 I-3; 項目 I-4）が他の質問項目と比べて低い値（ $M: 3.07$, $SD: 1.11$; $M: 3.14$, $SD: 1.20$; $M: 3.14$, $SD: 1.04$; $M: 3.21$, $SD: 1.06$ ）を示したことから、帯活動として英語特有の発音やイントネーションを学習する上でとりわけ自己効力感の低い学習者に対して効果が期待される（川井, 2015; 城一, 2016）、単語やフレーズ、文等をリズムに合わせて発音

練習をする「チャンツ」を取り入れた（河合，高山，2021）。また、自分や相手のこと、身の回りの物事について簡潔な表現を用いてその場で伝え合うことや質問し合うこと（表 2：項目 III-3）についても同様に低い自己効力感の値（ $M:3.34, SD:.99$ ）を示したことから、会話活動の難易度を段階的に高度化させた。具体的には、第 1 段階において身近な事柄について説明・理解することを、第 2 段階において多様な情報から内容を推測し合うことを、第 3 段階において課題解決に向けて議論することを目標とした。さらに、円滑な意思疎通と会話の維持と発展を助ける英語の応答表現も併せて導入した（表 3）。

表 3：授業内に実施した学習活動の詳細

第 1 段階：身近な事柄について説明・理解する	
発音練習	チャンツ（/an/, /at/, /ig/），リエゾン（音の連結・脱落・同化）
会話活動	自己・他己紹介する，物の配置について話す，ごみ出し規則の違いについて話す
応答表現	関心を示す，感情を表す，感想をたずねる
第 2 段階：多様な情報から内容を推測し合う	
発音練習	チャンツ（/en/, /ug/, /un/），リエゾン（音の連結・脱落・同化），英語の歌
会話活動	図形を用いて絵図を描き合う（クイズ），人や物を描写する（クイズ）
応答表現	相違の度合いを表す，考える時間を求める，共感を示す，同意を示す
第 3 段階：課題解決に向けた議論をする	
発音練習	チャンツ（/ox/, /in/, /ink/, /all/, /ing/），リエゾン（音の連結・脱落・同化）
会話活動	出来事を描写し合う（物語），予定を話し合う（交渉），社会課題について議論する（ミニ・ディベート）
応答表現	聞き直す，詳しい情報をたずねる，理解を示す，反意を示す，要点を確認する

2.5 解析方法

事前事後意識調査については、当該科目の初回授業時および最終授業時において受講者に対して実施した任意のアンケート調査結果の比較・分析（Mann-Whitney U 検定）と効果量の算出を行った。なお、有意性の判定は 10%未満とし、効果量の判定には Cohen（1988）による基準（ $d=.20$ ：効果量小， $d=.50$ ：効果量中， $d=.80$ ：効果量大）を用いた。また、記述データについては、コーディングを行ったのち各項目の回答頻度を求めた。さらに、事前事後の英語スピーキングテストについては、抽出者 4 名の発話録音を日本語母語話者の英語教師 2 名、英語母語話者の英語教師 2 名がそれぞれ、共通の評価シートを用いて判定を行い、その結果の分散分析を行った。なお、共通の評価シートは平成 26 年度に文部科学省が実施した英語力調査（文部科学省，2016）を参考に作成したものである。加えて、事後インタビューについては、録音データの逐語録をもとに英会話能力にかかわる記述を抽出し、そこからテーマや傾向を見出して分析を行った。

3. 結果

3.1 発音に関する評価

まず、対象者全体の発音に関する効力感についてであるが、事前事後意識調査の結果の分析（表 2）から、該当する 4 項目（表 2: 項目 I-1; 項目 I-2; 項目 I-3; 項目 I-4）いずれにおいても有意な差は示されなかった ($p<.10$)。

しかし、抽出者 4 名の事前事後の英語スピーキングテスト Part A の結果（表 4）から、音読による発音力について、分散分析の結果、学生 3 が有意に正の変化を示し ($F(1, 3) = \infty, p<.01$)、また、学生 2 も有意傾向の正の変化を示した ($F(1, 3) = 6.82, p<.10$)。この 2 名について、事前テストでは聞き手の理解に支障をきたす程度の発音の誤りや不明瞭さが観察されたが、事後テストでは発音の誤りが減少し流暢さが向上したと評価されている（表 6）。一方、有意な変化を示さなかった学生 1 および学生 4 については、評価のコメントから、すでに事前テストにおいて聞き手に理解可能な音読ができており事後テストにおいてもそれを維持していたことがうかがえる。加えて、意思疎通に差し支えないものの発音のミスや誤りが散見され、明らかな改善が見られないことも指摘されている（表 6）。

3.2 英会話力の向上に関する評価

続いて、英会話力の向上に関する評価であるが、対象者全体の話すことに関する効力感について、事前事後意識調査の結果の分析（表 2）から、該当する 6 項目（表 2: 項目 II-1; 項目 II-2; 項目 II-3; 項目 III-1; 項目 III-2; 項目 III-3）すべてにおいて有意な差は示されなかった ($p<.10$)。しかし、抽出者 4 名の事前事後の英語スピーキングテスト Part B および Part C の結果から（表 4）、学生 1、学生 2、学生 3、学生 4 それぞれ異なる評価項目において有意な変化が表れている。

まず、学生 1 についてであるが、即興を前提とするやりとり（Part B (1)）において、語彙や文法の運用能力が向上し ($F(1, 3) = 13.44, p<.05$)、表現の言い換えや的確な意思伝達ができるようになったことが評価されている（表 7）。

次に、学生 2 についてであるが、ある程度の準備をした上で話すこと（Part C (1)）において、その内容や構成面で向上し（ $F(1, 3)=25.00, p<.05$ ）、抽象的な事柄に具体例を加えて説明する等、聞き手に早く正確に伝える論理展開法が身についたことが特筆されている（表 9）。

表 4：事前事後スピーキングテスト評価の統計分析結果

Part A: 音読（発音・流暢さ）【0, 1, 2 点】													
	評価者 α		評価者 β		評価者 γ		評価者 δ		平均値 (M)		標準偏差 (SD)		F 比
	事前	事後	事前	事後	事前	事後	事前	事後	事前	事後	事前	事後	
学生 1	2	2	1	1.5	2	2	2	2	1.75	1.88	.43	.21	1.00 ns
学生 2	1	2	1	1.5	1	2	2	2	1.25	1.88	.43	.22	6.82 †
学生 3	0	1	0.5	1.5	1	2	1	2	.63	1.63	.41	.41	∞ **
学生 4	1	1	1.5	1	2	2	2	2	1.63	1.50	.41	.50	1.00 ns
Part B (1): 即興を前提とするやりとり（内容）【0, 1, 2, 3 点】													
	評価者 α		評価者 β		評価者 γ		評価者 δ		平均値 (M)		標準偏差 (SD)		F 比
	事前	事後	事前	事後	事前	事後	事前	事後	事前	事後	事前	事後	
学生 1	0	3	1	2	2	3	2	2	1.25	2.50	.83	.50	3.95 ns
学生 2	2	2	2	3	3	3	3	3	2.50	2.75	.50	.43	1.00 ns
学生 3	1	1	2	2	2	2	1	2	1.50	1.75	.50	.43	1.00 ns
学生 4	1	2	2	3	2	3	1	2	1.50	2.50	.50	.50	∞ **
Part B (2): 即興を前提とするやりとり（文法・表現）【0, 1, 2, 3 点】													
	評価者 α		評価者 β		評価者 γ		評価者 δ		平均値 (M)		標準偏差 (SD)		F 比
	事前	事後	事前	事後	事前	事後	事前	事後	事前	事後	事前	事後	
学生 1	0	2	1.5	2	2	3	1	3	1.13	2.50	.74	.50	13.44 *
学生 2	2	2	2	2.5	2	2	3	3	2.25	2.38	.43	.41	1.00 ns
学生 3	1	1	1	1.5	1	2	1	2	1.00	1.63	0	.41	6.82 †
学生 4	1	2	1.5	3	2	3	1	3	1.38	2.75	.41	.43	33.00 *
Part C (1): ある程度の準備をした上で話すこと（内容・構成）【0, 1, 2, 3 点】													
	評価者 α		評価者 β		評価者 γ		評価者 δ		平均値 (M)		標準偏差 (SD)		F 比
	事前	事後	事前	事後	事前	事後	事前	事後	事前	事後	事前	事後	
学生 1	3	3	2	3	2	2	3	2	2.50	2.50	.50	.50	.00 ns
学生 2	1	2	2	3	1	3	2	3	1.50	2.75	.50	.43	25.00 *
学生 3	0	3	0	3	0	2	0	3	0	2.75	0	.43	120.00 **
学生 4	2	3	3	3	2	2	3	3	2.50	2.75	.50	.43	1.00 ns
Part C (2): ある程度の準備をした上で話すこと（文法・表現）【0, 1, 2, 3 点】													
	評価者 α		評価者 β		評価者 γ		評価者 δ		平均値 (M)		標準偏差 (SD)		F 比
	事前	事後	事前	事後	事前	事後	事前	事後	事前	事後	事前	事後	
学生 1	3	3	1	2	2	3	3	3	2.25	2.75	.83	.43	3.00 ns
学生 2	2	3	1	3	0	2	3	3	1.50	2.75	1.12	.43	6.82 †
学生 3	1	3	0	3	1	2	0	3	0.50	2.75	.50	.43	22.09 *
学生 4	2	3	3	3	2	2	2	3	2.25	2.75	.43	.43	3.00 ns

† $p<.10$ * $p<.05$ ** $p<.01$

また、学生 3 についてであるが、ある程度の準備をした上で話すこと（Part C (1), Part C (2)）において、その内容や構成面での向上（ $F(1, 3)=120.00, p<.01$ ）に加え、語彙や文

法の運用能力も向上した ($F(1, 3)=22.09, p<.05$)。意思疎通に必要な語彙力を習得したことにより発問への理解力が大きく向上し、平易な語彙を駆使して早く正確に回答することができるようになった点が評価されている (表 8)。

そして、学生 4 についてであるが、即興を前提とするやりとり (Part B (1), Part B (2)) において、語彙や文法の運用能力についてより正確さが増すとともに ($F(1, 3)=33.00, p<.05$)、発問に対して不足無く適切な反応をしていることが評価されている (表 7)。他方、回答の論理性が損なわれている点や本文を直接引用している点も指摘されていたため、改めて録音を確認したところ、発問が直接引用を招くような内容であったこと、また、意見を求める発問については、本文の論理に必ずしも合致していないものの題材のテーマを発展的に捉えて考えが述べられており、論理性は保たれていた。

3.3 学習活動の有効性に関する調査結果

次に、授業において行われた多様な学習活動の有効性についてであるが、受講生を対象とした事後意識調査 ($N=28$) の分析から (表 5)、身近な社会問題について英語で討議する「ミニ・ディベート」の有効性を感じた学生が最も多く ($n=19$)、その他には「チャンツ」($n=5$)、留学生に対し大学生活をテーマごとに紹介する「プレゼンテーション」($n=2$)、その正確な名称を聞き出すために人物や事物、場所を描写する「クイズ」($n=1$)、グループ内で回し書きをしながら即興的に物語を創る「ストーリー・チェイン」($n=1$) が挙げられた。「ミニ・ディベート」、「クイズ」、「ストーリー・チェイン」はペアあるいはグループで行われ、インタラクションを伴う総合的な英語運用のできるグループ学習活動であった。それに対し、「チャンツ」は Google ドキュメントの音声入力機能を活用し、英語特有の音を含む英文を正しく入力されるまで繰り返し音読練習をする個別学習活動であった。

併せて、授業を通して向上した点については、「表現力」($n=16$) に分類される回答が最も多く、その他には「会話力」($n=7$)、「語彙力・文法力」($n=5$)、「発音」($n=2$)、「リスニング」($n=1$) に関する記述回答を得た (表 5)。なお、「表現力」には既に備わって

いる語彙力や文法力を駆使し、実際の英語スピーキング・コミュニケーションにおいて意思疎通を図ろうとする伝達力や、多様な相槌・応答表現を適切な文脈で用いる能力が含まれ、「ミニ・ディベート」で求められる総合的な英語運用能力と概ね重なる。

表 5：事後意識調査記述回答の総括

効果的と感じた学習活動 (N=28)		
活動の種類	活動の概要	回答数
クイズ	人や動物、物を描写し、それが何であるか当てる	1
ストーリー・チェイン	グループ内で物語を回し書きする	1
チャンツ	英語特有の音を含む英文を音読し、Google ドキュメントに正しく入力されているかどうか確認する	5
ミニ・ディベート	ある議題について肯定・否定の立場から討論する	19
プレゼンテーション	物語創作と読み聞かせ	2
授業を通して向上した点 (N=28)		
知識・能力	各項目の具体例	回答数
会話力	会話維持、会話意欲 (WTC)、適切な応答	7
語彙・文法力	イディオム、相槌・応答表現、単語	5
発音	発音意識、発音	2
表現力	語彙・文法の運用、伝達力、多様な相槌	16
リスニング	適切な理解	1

加えて、抽出者 4 名に対して実施した事後インタビューの結果についてであるが、学生 1、学生 2、学生 3、学生 4 とともに当該科目の授業が英会話能力向上に効果的であったと回答した上で、とりわけ「応答表現」の学習に有効性を感じていることが分かった。その理由として、話す時の表現の幅が広がり同じ表現を何度も使わなくなったこと、多様な表現があることを知り勉強になったことが言及された。また、「チャンツ」についても学生 2、学生 3、学生 4 の 3 名の学生が有効性を認めており、その理由として、英語らしい発音についても意識が向くようになったこと、学校教育現場で実際に使えることが挙げられた。しかし、「チャンツ」の詞を Google ドキュメントに音声入力することについて、人間と AI では音声の認識方法が異なるため正確に入力されるまで 20 回以上繰り返す必要がある点、さらに詞が子どもを対象とした拙い内容である点が課題として指摘された。その他には、毎授業の冒頭で行った「スモールトーク」および「自作の語彙ノート」に有効性を感じた学生が 1 名いた (学生 1)。その理由として、前者は出されたトピックについて即座に意見をまとめて話す力がつく点、後者は応答表現をはじめ授業で

学んだ語彙を文脈とともに書き残すことにより、新しい記憶を定着させることに役立つ点が言及された。さらに、「クイズ」および「ストーリー・チェイン」に有効性を感じた学生も1名いた（学生4）。その理由として、前者は相手意識をもって話すことができるようになった点、後者についても、他者に読まれることを意識し程良い緊張感をもって楽しみながらライティングに取り組むことができる点が挙げられた。

以上のことから、学習者が英会話能力の向上に役立つと感じる活動の特徴として、総合的英語運用および学習者間のインタラクションの機会が十分に設けられていること、また、文法的学習活動は応答表現等それに付随した内容であることが分かった。

3.4 授業外の学習活動に関する調査結果

しかし、教育効果の要因は必ずしも学科内の学習活動に限定されない可能性があることから、抽出者対象の事後インタビューにおいて、当該科目外の学習活動が学習者の英語スピーキング・コミュニケーション能力向上に与える影響を探った。その結果、学生2および学生3については、英語科指導法の授業で学んだ英語特有の発音に関する講義が発音への意識に影響していたことが分かった。例えば、/r/と/l/の発音の違いを紙の振幅で判別する方法を学び、それを自宅でも練習する等である。同様に、学生4も英語科教育に関連する授業において、ニュース記事や短い論文を読みそれについて自分の考えを英作文する等、英語に触れる機会が多くあり、さらに本研究対象の当該科目では英語で話す活動が多かったため、総合的に英語を用い相乗効果を感じていることが伺えた。

また、抽出者の学生4名ともに、自身の興味関心に合わせて大学の授業外で英語に触れていたことも分かった（表9）。学生1については、自宅で料理等の作業中に教育問題に関する海外の動画を日本語字幕付きで視聴し日常会話表現を学ぶ、学生2については、テレビのニュースやアニメ等の動画を英語音声で視聴し字幕を見てなるほどと思った英語表現を真似て言う、学生3については、好きなサッカーに関する英語のニュース記事を週2、3回の頻度で読む、英語の検定試験の勉強をする、学生4については、英語のドラマを日本語字幕付きで視聴する、毎日洋楽を流して聴く等が挙げられた。

表 6：事前事後スピーキングテスト Part A の評価コメント総括

Part A: 音読（発音・流暢さ）		
	事前	事後
学生 1	<ul style="list-style-type: none"> 強く明瞭なアメリカ英語である 1 音節を 2 音節で発音していた 語と語に間があり，抑揚や速度の変化に乏しく機械的である 速度が少し遅い 全体的に十分理解可能である 	<ul style="list-style-type: none"> 語と語の連結が改善した 発音のミスや読み誤りが少しある 流暢性に極めて優れている リエゾンや中間音ができていない 抑揚に欠ける 全体的に非常に理解しやすい
学生 2	<ul style="list-style-type: none"> 速度やイントネーションが良い 音節の長短が極端である 単数形と複数形が区別できていない リエゾンができていない 発音の誤りが多く理解できない 	<ul style="list-style-type: none"> 聞き取りやすくなったが，まだ改善の余地がある 全体的に流暢でよくできている 単語の読み誤りが少しあるが全体としてよくできている
学生 3	<ul style="list-style-type: none"> つぶやいているようで理解できない 発音が不明確である 子音のあとに母音が残る（母語の影響） 基本的な単語の発音に誤りがある 	<ul style="list-style-type: none"> 単数名詞を複数形で発音し続けている 全体的に単語の読み誤りが減少した 母語の影響が減少している 流暢さが向上している
学生 4	<ul style="list-style-type: none"> 速度やイントネーションが良い 抑揚やリズム，語強勢がとても良い /r/と/l/の区別，子音の発音が正確である いくつか単語の読み誤りがある 時に発音につまずきや詰まりがある 全体的に理解可能である 	<ul style="list-style-type: none"> 速すぎるところもあるが改善している 抑揚や強弱，リエゾンに優れている /s/と/sh/の区別ができていない 4 名中唯一 dove を正確に発音できた 全体的に良いが大きな変化はない

表 7：事前事後スピーキングテスト Part B の評価コメント総括

Part B (1): 即興を前提とするやりとり（内容）		
	事前	事後
学生 1	<ul style="list-style-type: none"> 質問を理解し，適切な回答をしている 	<ul style="list-style-type: none"> 発問に対しての反応が良くなっている 本文を直接引用して回答している
学生 2	<ul style="list-style-type: none"> 自信を持って回答していたが，説明が不足している 回答内容に矛盾がある 高い即興性を示している 全体として非常によくできている 	<ul style="list-style-type: none"> 十分理解可能な内容である 即興性に向上が見られる 本文の直接引用に解説を加えている
学生 3	<ul style="list-style-type: none"> 何が言いたいのか伝わらない 質問と回答内容が合っていない 直接引用せず自分の言葉で答えようとしている 	<ul style="list-style-type: none"> 何を言いたいのか分かりづらいが，大まかな部分は理解できる 発問の理解と反応が大きく向上している 本文を直接引用して回答している 2 つ回答があるうち 1 つは述べていた
学生 4	<ul style="list-style-type: none"> 回答をすべて述べないで終えている 本文を直接引用して回答している 	<ul style="list-style-type: none"> 自信が伝わる十分理解可能な回答である 回答の論理性が損なわれている 発問に対して適切な反応している 本文を直接引用して回答している
Part B (2): 即興を前提とするやりとり（文法・表現）		
	事前	事後
学生 1	<ul style="list-style-type: none"> 本文中の表現を用いて回答している 自分の言葉で話すことはできない 	<ul style="list-style-type: none"> 受動態の使い方が誤っていた 本文中の表現の言い換えができていた 聞き手の理解をほとんど妨げない
学生 2	<ul style="list-style-type: none"> 言葉に詰まることがない 語順に誤りがある 途切れがあり理解しづらい 	<ul style="list-style-type: none"> 細かな文法の誤りがあるが，著しく理解を妨げるものではない
学生 3	<ul style="list-style-type: none"> 基本的な語彙が使われている 語順・品詞の用法に誤りがある 語彙や文法の知識が不足している 即興性はある程度認められる 	<ul style="list-style-type: none"> 適切な語彙選択ができていない 語句の使用に文法的誤りがある
学生 4	<ul style="list-style-type: none"> すばやく明確に回答している キーワードの発話が伝わりづらい 	<ul style="list-style-type: none"> 適切な語彙を用いて回答できている 適切な文法を用いて回答できている

表 8：事前事後スピーキングテスト Part C の評価コメント総括

Part C (1): ある程度の準備をした上で話すこと（内容・構成）		
	事前	事後
学生 1	<ul style="list-style-type: none"> 英語で発話内容を構築できていない 無言が続き、回答できていない 	<ul style="list-style-type: none"> すべての質問に正確に回答できている 英語で話すことへの抵抗感が減った 発問に対しての反応が良くなった 自分の考えをまとめることが課題
学生 2	<ul style="list-style-type: none"> 自分の言葉で話そうとしている 物語の意図と逆の内容が伝わる可能性 	<ul style="list-style-type: none"> 抽象的な論に例を挙げて説明していた 会話における論理展開法が定着している 質問や本文の意図を理解する力が向上している 発問への反応が早く正確であった
学生 3	<ul style="list-style-type: none"> 論理性やまとまりに欠ける回答であった 意味を成す回答を構築できていない 質問をしながら解決しようとしている 思い浮かんだことを発話している 	<ul style="list-style-type: none"> すべての発問に正確に回答できている 発問への反応が早く、理解力も大きく向上している
学生 4	<ul style="list-style-type: none"> 前置きの英語表現を用いながら考えを整理する手順が身につけている 質問を理解した上で即興性も認められた 	<ul style="list-style-type: none"> すべての発問に正確に回答できている 冒頭で自分の考えを明らかにしている 最後に自分の考えをまとめている 対話の基本型を心得ている
Part C (2): ある程度の準備をした上で話すこと（文法・表現）		
	事前	事後
学生 1	<ul style="list-style-type: none"> 文法の誤りが理解の妨げになる 自分の考えを表現する語彙が不足している 	<ul style="list-style-type: none"> 長い回答を即興で発することができる 回答に詰まると聞き手の援助が必要
学生 2	<ul style="list-style-type: none"> 文法、語順に誤りがあった 発音や速さについては問題ない 	<ul style="list-style-type: none"> 初歩的な語彙が用いられている 有意な文を即興で構築できている
学生 3	<ul style="list-style-type: none"> ところどころ適切な語彙が使われていた 意思疎通に必要な語彙力が不足している 	<ul style="list-style-type: none"> 平易な語彙を用いて回答ができている
学生 4	<ul style="list-style-type: none"> 語彙や文構築の面で課題がある 聞き取りづらい箇所があった 	<ul style="list-style-type: none"> 速さ、発音、抑揚ともに素晴らしい 途切れがあり文単位の抑揚が理解しづらい

表 9：事後インタビューでの授業外学習活動に関する回答の総括

授業外の英語学習活動		
	他科目に関する取組	個人的な取組
学生 1	<ul style="list-style-type: none"> 特になし 	<ul style="list-style-type: none"> 教育問題に関連する話題を扱っている英語母語話者による配信動画を日本語字幕付き英語音声で週に 1 回の頻度で視聴する
学生 2	<ul style="list-style-type: none"> 言及なし 	<ul style="list-style-type: none"> テレビのニュースやアニメ等の動画を日本語字幕付き英語音声で視聴する
学生 3	<ul style="list-style-type: none"> 自宅で発音の練習する 英語特有の発音に関する講義を受けた 	<ul style="list-style-type: none"> サッカー関連の英語ニュース記事を週 2、3 回の頻度で読む 英語の検定試験の勉強をする
学生 4	<ul style="list-style-type: none"> 授業の課題としてニュース記事や短い論文を読みそれについて自分の考えを英作文する 授業の課題として英語で授業の導入方法を考える 	<ul style="list-style-type: none"> 英語のドラマを日本語の字幕付き英語音声で視聴する 毎日洋楽を流して聴く

4. 考察

4.1 多様な学習活動の取捨選択とその有効性

本研究では、事前意識調査を通して受講者のニーズを把握し、それに沿って学習活動の取捨選択を行った。当該科目については、「発音」、「身近な事柄について話し合う」とに学習者の低い効力感が見られたため、これらを克服することを念頭にコースを展開させていった。「発音」については帯活動として「チャンツ」の学習活動を継続的に行い、

「身近なことについて話し合う」については、使用テキスト（Wicking 他, 2022）にあるタスクを取捨選択、さらに受講者が取り組みやすい内容に修正や追加、差し替えを行った。このように、学習者と折り合いをつけながらコース内容を編み上げていくプロセス・シラバス（Breen & Littlejohn, 2000）の要素を取り入れ、テキストの各課を始めから終わりまで順番に進めることはせずに、3 つの段階（表 3）の範囲内で学習者のニーズや傾向、習熟度に合わせて各段階の繋がりを維持しつつコースを展開させた。以下にこれら学習活動の有効性および課題について考察を述べる。

まず、「チャンツ」の有効性についてであるが、対象者全体としては「発音」の効力感に有意な変化が見られなかった。しかし、抽出者については事後テストの評価が有意に向上した学生もあり、最終的に全員が聞き手に理解可能な発音をできるようになった。なお、対象者全体に対する事後意識調査においても「チャンツ」を効果的であると回答した学生が 5 名おり、流暢さを高めることに有効性が認められた。だが、事後インタビューでも指摘された通り、「チャンツ」の詞が小学生向けの拙い内容であったため、大人の学習者の知的水準に配慮し、今後、高等教育機関における教養教育にも配慮しマザー・グースや英詩等を題材にする可能性を検討していく。

次に、「身近なことについて話し合う」については、対象者全体としては効力感に有意な変化は見られなかったが、抽出者については事後テストにおいて全員が発問に対し適切な応答ができるようになり、文法、表現、内容、構成のいずれかあるいは複数の領域において評価が有意に向上した。また、対象者全体の事後意識調査から「表現力」が身についたと回答している学生が 16 名いたことから、「クイズ」や「ストーリー・チェイン」、「ミニ・ディベート」等、学習者のニーズに合わせた多様な学習活動を段階的に系統立てて導入することについても一定の有効性が認められた。他方、事後の英語スピーキングテストでも観察された通り、会話の内容について論理性に欠ける等、未克服の課題も残った。今後の学習活動をより充実させるために、各学習者の発話に注意深く耳を傾け、ニーズに即した段階的指導や効果的なフィードバック、振り返りを行うことが改善点として挙げられる。

4.2 相槌・応答表現学習の意義と課題

「相槌・応答表現」の学習活動について、受講者に対する事後意識調査の結果および抽出者に対する事後インタビューの結果から、有効性を感じている学生が多いことが分かった。相手の発話に対し何らかの反応を示すことで自身の理解や考えを表し、会話の維持や進展に繋がる、円滑な意思疎通を図る上で等閑視できない学習項目である（岩田, 2023）。さらに、ペアやグループでのインタラクション（やりとり）を通して英語を話すことへの不安が克服されるだけでなく、英会話能力を向上させていく上で必要な気づきを得ることも期待される（Long, 1996）。実際、受講者全員に対して実施した事後意識調査において、英会話能力を向上させていく上で効果的な学習活動として最も多くの回答を得たのは「ミニ・ディベート」であったが、付随的に学習した相槌・応答表現をディベートにおいて用いることで学習者は円滑な意思疎通を維持することができ、インタラクション（やりとり）を通して学びや気づきを得る機会が多かったことが、その有効性の高さの一要因と考えられる。

「相槌・応答表現」は、5 語以内の英単語からなる平易な英語表現であることが多いが、抽出者の 1 人である学生 2 が事後インタビューにおいて、当該科目を通して応答表現を学習するまでは“*That’s nice.*”しか使えなかったと述べていたことから、「表現力」が向上したと感じた受講生の多くはこれまでの英語学習歴において「相槌・応答表現」等の英会話において使用される英語表現に触れる機会が少なかったことが推察される（柳田, 2014）。「相槌・応答表現」は多種多様で分類することが難しく、また、その使用は場面や状況、話者の選択に依拠する側面があり、英文法事項のように明瞭に教授することが難しい（Roever, 2022）。本研究の授業実践においても、想定する場面で使用される可能性のある「相槌・応答表現」は複数に及び、適切な分量の範囲内での選択や用例の提示に迷うことが多くあった。言語運用と密接にかかわりのある、学習者の社会言語的能力や方略的能力（Canale & Swain, 1980）の向上を期待できる教授方法についてさらに研究を行っていく必要がある。

4.3. 授業外の学習活動から得る教育的示唆

本研究では、抽出者 4 名に対する事後インタビューから、学生 1、学生 2、学生 3、学生 4 とともに自身の目的や興味関心に合わせて当該科目の授業外で英語に触れていることが分かった。いずれの学生もインターネットやテレビを活用して日本語字幕付き英語音声で動画の視聴やニュース記事の購読を週 1 回から 3 回程度の頻度で行っていた。当該学生らは、これらの学習活動を勉強と捉えておらず、実際に使われている英語、例えば授業で習った「相槌・応答表現」が自然な会話中でどのように使われているか、他にどのような表現があるのかを学ぶ機会であるという認識を示していた。ある場面や状況において用いられるさまざまな英語表現を動画等で例示することは、授業内においても可能であり、今後、その実践効果をはかりたい。また、学生 3、学生 4 については英語教員免許取得に必要な科目の授業の復習や課題に取り組むことで英語学習により多くの時間を費やしていることや、教える立場から英語学習を見る視点を得ていることも分かった。そして学生 4 については、まとまった文章を読むことや書くことが自分の意見を頭の中でまとめて話すことに役立ったと述べていたことから、英会話能力の向上を目指すコースにおいても多様な学習傾向に配慮した学習形態の活動を織り交ぜることの重要性を再確認した。例えば、当該授業の場合、発話の論理性や一貫性を向上させることを目的とし、「ミニ・ディベート」の準備学習および事後学習としてリーディング活動やライティング活動も統合的に取り入れることができると考える。

4.4. まとめ

以上、本研究の調査の結果から、学習者のニーズに応じて多様な学習活動を取捨選択しそれらを段階的に導入した教育実践の効果について、対象者全体の意識調査からは有意な変化は見られなかったが、抽出者 4 名を対象とした英語スピーキングテストの評価からは有意な向上が観察された。4 名ともに理解可能な発話による意思疎通を図る総合的な英語運用能力を向上させることができた。そしてこのことから、津田 (p.125) が示唆する通り、学習者の自己効力感が必ずしも実際の英語能力と合致しない可能性がある

ことが確認された。それと同時に、授業内で行ったさまざまな学習活動の中でも、付随的に導入した「相槌・応答表現」の学習活動に有効性を感じている学習者が多いことも分かった。円滑な言語コミュニケーションを図る上で等閑視できない「相槌・応答表現」であるが、場面や状況において話者の選択によるところが多く、多種多様で文脈により流動的であるため、今後、何をどのように教授することができるのか、学習者の授業外における学習活動状況にも鑑みてさらに具体的な指導研究を進めていく必要がある。

最後に、今後の課題について述べる。本研究では多様な学習活動の効果を対象者全体の効力感および抽出者 4 名に対する英語スピーキングテストの評価によってはかったが、学習者の効力感の変化をより正確にはかるには 3 ヶ月以上の研究期間が必要であり、さらに、発音に関する効力感について、自己判断ではかることに限界もあると考えられる。より長期的な研究を行うとともに、測定方法も、先行研究における効力感尺度を参照し妥当性を再検討する必要がある。さらに、各学習活動の効果についても、学習者の変容とあわせて分析することで個別最適な学習支援に繋がる示唆を得ることが期待される。加えて、本研究の事前事後に実施した英語スピーキングテストを評価するにあたり、評価者から評価の基準や観点について確認のための質問があった。共通理解を得るために、評価シートと併せて補足説明を行う必要性が明らかになった。また、評価の幅も Part A は「0, 1, 2」、Part B は「0, 1, 2, 3」と異なる尺度を採用したことで、分析結果の比較において不明瞭さを招いた。これらのことを踏まえ、試行テストの反省点を重んじるとともに、先行研究に習い信頼性と妥当性に留意したテストと評価を再検討していく。

参考文献

イーオン (2021). 『小学校の英語教育に関する教員意識調査 2021』

https://www.aeonet.co.jp/company/information/newsrelease/pdf/aeon_210315.pdf (2024 年 11 月 7 日参照)

岩田裕子 (2023 年 8 月 31 日). 語用指導の一試案－高校の検定教科書を用いた会話指導方法の課題と可能性－ [学会発表]. 第 62 回大学英語教育学会大会, 東京, 日本.

- 城一道子 (2016). 「フォニックス・ライム・チャンツ・歌を活用した発音指導の教育効果：TAE (Thinking at the Edge)を使用した分析」『教育総合研究：江戸川大学教職課程センター紀要』, 4, 1-12.
- 川井一枝 (2015). 「英語発音指導におけるチャンツの有効性」『いわき明星大学人文学部研究紀要』, 28, 199-210.
- 河合裕美, 高山芳樹 (2021). 『きいて・みて・まねて覚える英語の音 動画でできる音声指導』大修館.
- 文部科学省 (2016). 『英語に関する論点 (案)』
https://www.mext.go.jp/b_menu/shingi/chousa/shotou/112/shiryo/__icsFiles/afieldfile/2016/01/20/1364552_1.pdf (2024 年 11 月 7 日参照)
- 文部科学省 (2017). 『小学校外国語活動・外国語研修ガイドブック』
https://www.mext.go.jp/a_menu/kokusai/gaikokugo/__icsFiles/afieldfile/2017/07/07/1387503_1.pdf (2024 年 11 月 7 日参照)
- 文部科学省 (2018). 『外国語活動・外国語の目標の学校段階別一覧表』
https://www.mext.go.jp/content/1407196_26_1.pdf (2024 年 11 月 7 日参照)
- 文部科学省 (2019a). 『外国語 (英語) コアカリキュラム』
https://www.mext.go.jp/component/a_menu/education/detail/__icsFiles/afieldfile/2019/04/04/1415122_3.pdf (2024 年 11 月 7 日参照)
- 文部科学省 (2019b). 『小学校学習指導要領 (平成 29 年告示) 解説 (外国語活動・外国語編)』
https://www.mext.go.jp/component/a_menu/education/micro_detail/icsFiles/afieldfile/2019/03/18/1387017_011.pdf (2024 年 11 月 7 日参照)
- 田中真紀子 (2023 年 1 月 31 日). 「小学校英語の課題 教員『発音』に不安強く」『日本経済新聞』
https://www.kandagaigo.ac.jp/group_news/data/9515/block/file_230131170533000000_0_1.pdf (2024 年 11 月 7 日参照)

- 津田敦子 (2022). 「教員養成課程の学生の小学校英語に対する不安の分析」『琉球大学教育学部紀要』, 101, 119–129.
- 柳田恵美子 (2014). 「コミュニケーション能力とは何か」上智大学 CLT プロジェクト編『コミュニケーション能力を高める英語教育を考える』(pp.88-91). アルク出版.
- 米崎里, 多良静也, 佃由紀子 (2016). 「小学校外国語活動の教科化・低学年化に対する小学校教員の不安—その構造と変遷—」『小学校英語教育学会誌』, 16(1), 132–146.
- Wicking, P., 田村祐, 加藤由崇, 小林真実, 松村昌紀, 横山友里 (2022). 『Getting Things Done [Book 1]—Tasks for Connecting the Classroom with the Real World』三修社.
- Breen, M. P. & Littlejohn, A. (2000). *Classroom decision-making: Negotiation and process syllabuses in practice*. Cambridge University Press.
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied linguistics*, 1(1), 1-47.
- Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Routledge.
- Jones, V. V. S. (1912) *Aesop's Fables*. The Project Gutenberg eBook of AESOP'S FABLES. Retrieved November 7, 2024, from <https://www.gutenberg.org/files/11339/11339-h/11339-h.htm>
- Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. Ritchie and T. Bhatia (eds). *Handbook of Second Language Acquisition*. Academic Press, 2, 413-468.
- Roever, C. (2022). *Teaching and testing second language pragmatics and interaction*. Routledge.

謝辞

瀧澤英子氏、中野雅也氏、Shaun Harris 氏、Roberto Carlos Luna 氏には、本研究の英語スピーキングテスト評価について有益な助言をいただいた。

DeepL Write を活用した英語ライティング指導と訂正フィードバック

Instruction and Feedback in English Writing Using DeepL Write

甲斐 順（帝京大学）

キーワード：DeepL Write、英語ライティング、訂正フィードバック、アンケート

Abstract

This study investigates Japanese university students' perceptions of using DeepL Write for English essay writing and receiving written corrective feedback (WCF) via the AI tool. Out of thirty-four sophomores enrolled in a general English course, twenty-eight responded to a questionnaire on using AI tools for writing, including their feelings about receiving WCF from the tool. The findings indicate that while many students were unfamiliar with DeepL Write prior to the lesson, a majority appreciated the feedback provided by the technology, and most of them expressed interest in using it for future writing tasks.

1. はじめに

英語ライティング指導において英語教員の悩みどころは、どの程度訂正フィードバックを学習者に行えばよいかということである。学習者の誤りはすべて直さなければならないという教員の信条があるため、必要以上の重労働に陥る可能性がある(佐野, 2013)。

近年、Grammarly や DeepL Write などの AI (Artificial Intelligence)駆動型の文章サポートツールや ChatGPT や Copilot などの生成系 AI が普及し、英語ライティングの分野において威力を発揮しつつある。ただし、これらは進化の途上にあり、使用するツールにより精度が異なる(甲斐, 2024)ことを理解した上で、訂正フィードバックを行う必要がある。

本稿では、英語ライティング指導の中で、DeepL Write を活用し、訂正フィードバック

を行うことで、英語学習者の意識を明らかにすることを目的として行う。

2. 先行研究

第二言語ライティングにおける訂正フィードバックについては、効果がなく、有害であるなどとして廃止すべきとする Truscott (1996)の主張に対して、彼の研究の不十分さなどを指摘し反論した Ferris (1999)に、Truscott (1999)が再反論するといった形で、他の研究者も論争に加わり関心が高まっていく。その後、両者が求めていた実証研究が報告され、訂正フィードバックにより学習者の正確性の向上やその効果が持続することが明らかになっていく (Hyland, Nicolás-Conesa, and Cerezo, 2018)。そして、提出したライティング課題に訂正とコメントを求める学習者からの要求には応えるという姿勢が広まる (佐野, 2023)。ただし、教員が訂正フィードバックやコメントを書く作業には時間と労力がかかるため、コンピュータによる自動評価研究が行われるようになる (同上, 2023)。

近年、日本人英語学習者の英語ライティングに、AI 駆動型の支援ツールを利用した研究が報告されている。新美・梅木(2024)では、大学生に Grammarly を用いて英語ライティングの修正活動を行わせた結果、文法や表記の面で明らかな伸長が見られた。Gold (2023)は、成人 1 名を対象に Grammarly と DeepL を用いたライティングの効果を検証し、これらは重要なアフォーダンスを提供し、学習の足場掛けとなる支援を行い、認知的負担や不安を低下させ、ライティングに対する自信や動機を高めると報告している。山下・山中・杉山(2024)は、Grammarly、DeepL、ChatGPT の複数の AI ツールを搭載した Transable を大学生に使わせ、自動採点・評価の点で課題はあるが、自動添削の点で有用性が確認され、学生が自動添削・自動評価機能に満足していると報告している。

3. 研究の目的

AI ツールを英語ライティングで活用することに対して英語学習者はどのように受けとめているのだろうか。本研究では、AI による文章構成アシストツールである DeepL Write を学習者に使用させて作文することを認めるとともに、教員からの訂正フィード

バックに DeepL Write を使用し、学習者の意識を調査することを目的として行う。そこで、次の 9 つの研究課題を設定した。

研究課題 1 学習者は英語学習で AI ツールを使用したことがあるか。あればどのようなツールを使用しているか。

研究課題 2 学習者は授業受講以前に DeepL Write を知っていたか。

研究課題 3 学習者は DeepL Write の修正案についてどのように感じているか。

研究課題 4 学習者は授業内でどの程度 DeepL Write を利用していたか。

研究課題 5 授業内で DeepL Write を利用した学習者は、どのような感想を持っているか。また、継続して利用する意思があるか。

研究課題 6 授業内で DeepL Write を利用しなかった学習者はどのような理由で利用しなかったのか。

研究課題 7 授業で AI ツールを利用して英作文を行うことについてどう思っているか。

研究課題 8 学習者は授業で取り組んだ英作文について楽しいと感じたか。

研究課題 9 学習者は授業を通じて書く力がついたと感じたか。

4. DeepL Write について

DeepL Write は AI 搭載の文章作成アシストツールで、文法やスペリングなどの修正や別な言い回しの提案をとおり、文章作成を支援する(DeepL, 2024)。英語については、イギリス英語とアメリカ英語などに対応している(同上)。DeepL Write 左側の入力欄にテキストを入力すると、すぐに右側の出力欄に改善されたテキストが表示され、変更は緑色で表示される(同上)。DeepL Write は無料で利用が可能で、携帯電話上でも利用できることから授業内で学生に使用させることとした。

5. 実践

5.1 対象者

日本の私立大学で英語を専攻としない文系学部に所属する大学 2 年生で、筆者が指導

する前期の一般英語クラスの 34 人を調査対象者とした。英検の取得状況は 2 級が 8 人、準 2 級が 12 人、3 級が 5 人、英検未取得者が 9 人であった。なお、調査対象者となる学生からは、本研究の目的を説明し、協力への了承を得ている。

5.2 実践内容

一般英語クラスの使用教科書は『もっと知りたい！社会のいま 2 Trend Watching 2』（成美堂）である。全 20 章で各章は、Pre-Reading Vocabulary Task、Reading、Comprehension Questions、Vocabulary Practice、Read and Listen、Writing からなる。前期授業では、2 週で 1 章を終えるようにし、上記タスクの順で実施した。

ライティングの活動は、前期授業 15 回のうち、6 回実施した。教科書の Writing の自由作文に関する設問を基本的に活用し、各章が終わるたびに組み合わせた。参考までに、Chapter 1 Fake Online Reviews の設問は“Do you use reviews and ratings when you make an online purchase? Explain your answer.”である。

1 コマの授業時間は 1 時間 30 分で、ライティング実施日は、授業終了前の 15～20 分をこの活動に当てた。1 回目と 6 回目は携帯電話等を使わずに書くように指示し、2～5 回は DeepL Write を使用しても構わないことを伝え、組み合わせた。

DeepL Write の使い方については、まず 1 回目の自由英作文を返却後、DeepL Write を紹介した。DeepL Write については、①AI を利用した無料で使えるツールであること、②パソコンや携帯電話で使用可能であること、③英文を作成中に綴りや文法等の修正候補を緑色で示してくれる便利な支援ツールであること、の 3 つの利点を説明した。ただし、「1 回目の作文で true と書きたかった学生が ture と書いていたにもかかわらず、DeepL Write は修正候補すら示さず、直していなかった」という事例、及び “I couldn’t go to school yesterday. Because I was sick in bed.”と書いても DeepL Write はこの Because を見落とす」といった事例を紹介し、使用に関して注意するように説明した。そして次週の 2 回目の英作文に取り組む前に、もう 1 度 DeepL Write について簡単に説明した後、携帯電話の検索サイト（Google 等）で DeepL Write と打ち込み、まずサイトにたどり着

くよう促した。そして、DeepL Write のサイトに到達後、言語をイギリス英語からアメリカ英語に変える操作手順（図 1）及び、テキストを入力するとどのような修正案が提示されるかを示した図（図 2）をそれぞれスライドで提示した。



図 1. 携帯電話における DeepL Write の言語選択の操作手順スライド

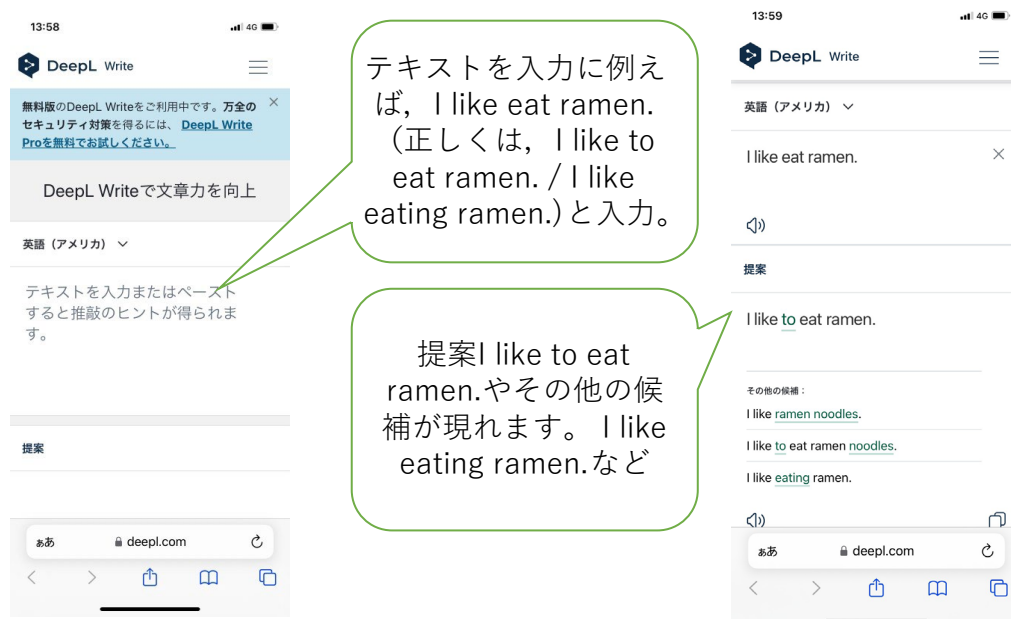


図 2. DeepL Write のテキスト入力方法と修正案のスライド

説明後の DeepL Write の使用については、学生に任せたが、特に使用方法に関して質問はなかった。なお、DeepL Write を使った場合は、使用回数を明記した項目（「使わなかった」、「1 回使った」、「2 回使った」、「3 回使った」、「4 回以上使った」）から選ばせた上で、提出させた。

学生が提出した作文は、マイクロソフト・ワードのディクテーション機能を活用して筆者が読み上げて文字起こしをした。そして、学生が書いた原文と相違ないことを確認し、ワードの文字カウントを使って、語数を記入しワード文書に保存した。各自が書いた作文を DeepL Write の入力欄に貼り付けると、瞬時に修正されたテキストが出力欄に表示されるので、プリントスクリーン機能を用いて複写し、別のワード文書に貼り付け全員分保存した。図 3 は DeepL Write を 2 回使用した学生の原文（左側）と出力結果（右側）である。

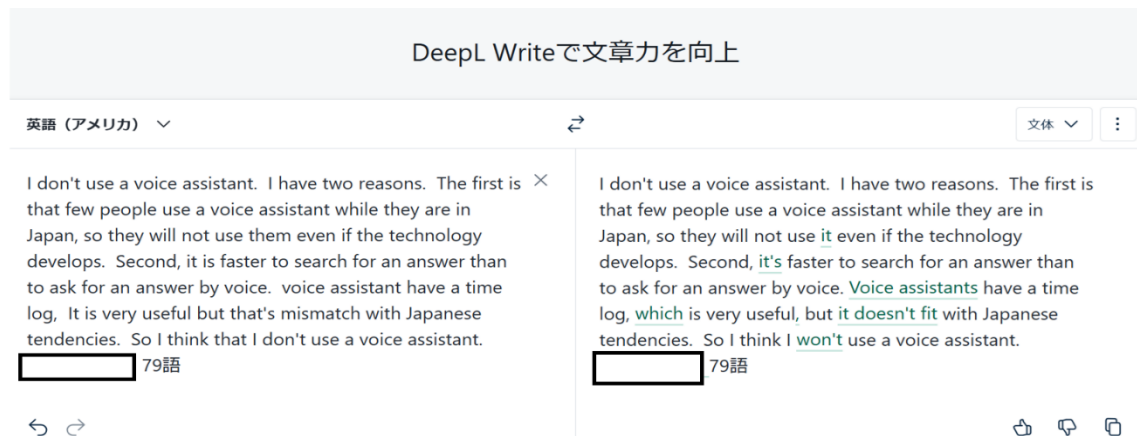


図 3. DeepL Write 画面の学生作文の入力と出力結果

DeepL Write が未修正の箇所については、元原稿に筆者が修正を加え、コメントも添えた。翌週の授業で、学生に元原稿と図 3 の用紙の合計 2 枚を返却し、修正部分を確認させた。その後、パワーポイントでよく書けている作品を学生の名前を伏せて紹介した。これは、学習者の作文の中から模範となる作文を選び、クラスに紹介することがあられ

るレベルの学習者にとって有意義であると思われるからである(鈴木, 2017)。全体に共通する誤りについてはスライドで示し簡単な解説を行った。

5.3 データ収集と分析方法

前期 14 回目の授業時にライティング活動に関するアンケートを実施した。回答にあたって、成績評価とは一切関係ないこと、無記名であること、回答するかどうかは自由参加であることを伝え、できるだけ協力するように依頼した。

アンケートは 12 の設問で構成されていた。設問 1 は、「この授業を受ける以前に、英語学習のために AI ツールを使用したことがありますか。」で、「ある」、「ない」、「わからない」の 3 択から選択する設問であった。「ある」と回答した場合は、設問 2「使用したことのある AI ツールを次の中から選んでください。」で、11 の選択肢（Google 翻訳、DeepL、DeepL Write、Grammarly、Ginger、ChatGPT、Copilot、Gemini、Siri、Alexa、その他）から複数回答が可能な設問となっていた。

設問 3「この授業で取り組んだ英作文について楽しかったですか。」、設問 4「この授業で取り組んだ英作文を通じて英語を書く力はついたと思いますか。」は 5 件法（「とても当てはまる」、「やや当てはまる」、「どちらともいえない」、「あまり当てはまらない」、「全く当てはまらない」）で回答することになっていた。

設問 5 は「あなたはこの授業を受ける以前に DeepL Write を知っていましたか。」に対して、2 件法（「知っていた」、「知らなかった」）に回答する設問、設問 6 は「自分の書いた英文と DeepL Write の修正案を見比べてどのような印象を持ちましたか。」に対して自由に記述する設問となっていた。

設問 7 は「この授業では DeepL Write を利用して英作文を書くことができました (Chapters 2-5)。あなたは DeepL Write を利用していましたか。」について、5 件法（「毎回利用した」、「ほぼ毎回利用した」、「初めは利用したが、だんだん利用しなくなった」、「あまり利用しなかった」、「全く利用しなかった」）に回答する設問であった。設問 8 は設問 7 で「利用した」と回答した学生が回答する設問「DeepL Write を利用してどのような

感想を持ちましたか。」で自由に記述してもらった。設問 9 も同じ学生が回答する設間で

「これからも引き続き DeepL Write を利用したいと思いますか。」について、5 件法（「とてもそう思う」、「ややそう思う」、「どちらでもない」、「あまりそう思わない」、「全くそう思わない」）で回答する設間であった。さらに設問 10 でその理由を書いてもらった。

設問 11 は設問 7 で「利用しなかった」と回答した学生に DeepL Write を利用しなかった理由をたずねる設間でこちらも自由に記述してもらった。

最後の設問 12 は「授業で AI ツールを利用して作文を行うことについてどう思いますか。」に対して自由記述で回答してもらった。

自由記述以外の設間については単純集計し、必要に応じてグラフ化した。自由記述の設間に対する回答については、似たような回答が複数含まれている場合、カテゴリーに分類し、分析を行った。

6. 結果

無記名のアンケートに回答したのは 28 人であった。アンケートの設問 1 は、授業受講以前の英語学習における AI ツールの使用の有無に関してであった。「ある」と回答した学生が 24 人、「ない」が 2 人、「どちらともいえない」が 2 人で、85%以上が受講以前に使用したと回答していた。

設問 2 は設問 1 で「ある」と回答した学生が答える設間で、使用したことのある AI ツールを 11 の選択肢から複数選べるものであった。図 4 はその結果を表したものである。

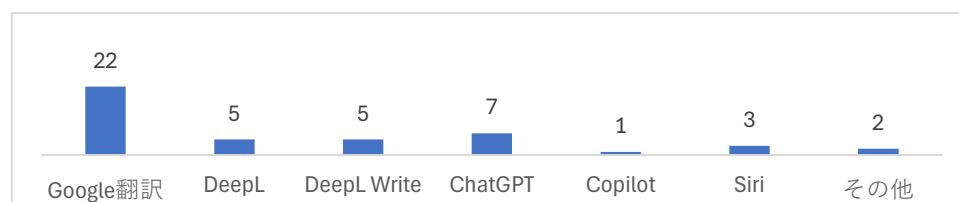


図 4. 使用したことのある AI ツール

Google 翻訳が 22 人で最も多く利用し、続いて ChatGPT が 7 人、DeepL と DeepL Write

を 5 人、Siri が 3 人で、その他が 2 人(Line が 1 人、iPhone のアプリが 1 人)であった。

設問 3 は授業で取り組んだ英作文について楽しかったかどうかをたずねる設問であった。図 5 はその結果を表したものである。

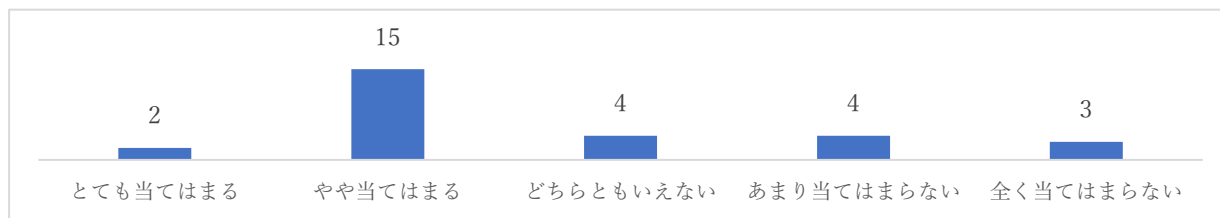


図 5. 授業で取り組んだ英作文についての楽しさ(n=28)

結果は「とても当てはまる」と「やや当てはまる」を合わせて 17 人で、約 60%の学生が英作文を楽しかったと実感していた。一方、「あまり当てはまらない」と「全く当てはまらない」の合計が 7 人で 25%の学生は英作文を楽しいとはとらえていなかった。残る 4 人は「どちらともいえない」と回答した。

設問 4 は授業で取り組んだ英作文を通じて英語を書く力がついたかどうかに関する設問で、図 6 はその結果を表したものである。

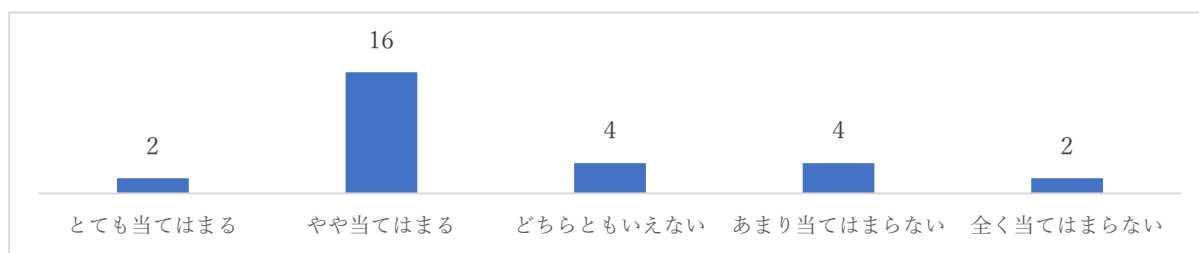


図 6. 英作文を通じて英語力がついたという実感(n=28)

結果は「とても当てはまる」と「やや当てはまる」を合わせて 18 人で、約 64%の学生が書く力がついたと実感していた。一方、「あまり当てはまらない」と「全く当てはまらない」の合計が 6 人で約 21%の学生はそう実感していなかったことが分かる。残りの 4

人が「どちらともいえない」と回答していた。

設問 5 は授業を受ける以前に DeepL Write を知っていたかどうかたずねるもので、「知っていた」の 7 人に対して、「知らなかった」が 21 人で、4 分の 3 の学生が DeepL Write については未知であったことが分かる。

設問 6 は自分が書いた英文と DeepL Write の修正案を見比べた時の印象を自由記述で回答するものであった。記述から次の 5 つのカテゴリーに分類することができた。

(1) 文法力の欠如や基礎的な間違いの自覚

DeepL Write の修正を通じて、自身の文法力の不足や綴りの誤りなどの基礎的な間違いを認識する学生が多く見られた。以下に 2 例ほど自由記述を示す。

- ・文法がまだ完全に身につけていないことがはっきり目に見えたので勉強が足りないと感じました。
- ・ to や with カンマ ピリオドなど細かい部分がぬけていることが多かった。また、スペルミスもあった。

(2) 文章の洗練度・表現の豊かさへの気づき

文がより洗練されたり、大人びた表現に変わったと感じる学生が多く見られた。また、自分の文章が単純であることを実感し、表現の幅を広げる必要性を感じている。さらに、伝わりやすさや自然な言い回しを学ぶきっかけになっている回答も見られた。次の具体例の 1 つ目は洗練されたことを 2 つ目は大人びた印象を表している。

- ・自分が書いたものよりもきれいな文章になっていると思いました。
- ・自分の英文は単純であり、中学生で習うような口調であったが、DeepL Write の修正案からは同じ内容でもより大人びた印象を受けた。

(3) 語彙力の不足を痛感

自分の語彙力が不足していると感じる学生も一定数おり、より表現豊かな文章を書くために、語彙を増やす必要性を感じていた。このカテゴリーに当てはまる例を示す。

- ・自分の知らない単語がたくさん出てきて、単語不足を実感した。
- ・全体的に語彙力が足りていないことがわかったため、語彙を増やしていかなければ

ばならないと感じた。

(4) 学習ツールとしての有用性の実感

DeepL Write の指摘が細かく、具体的なミスを発見できる点を学習に役立つと感じている学生がいる。単なる修正ツールではなく、自学習の補助ツールとして活用できると考えている。具体例として 2 つのコメントを紹介しておきたい。

- ・正しい修正点が見やすく、自分で探して何に直すかを考える作業が省け分かりやすかった。
- ・知識が深まり、学習ツールとしてとても使える印象だった。

(5) 修正結果に対する驚きや疑問

予想していた使い方とは異なる修正がされて驚いたり、納得できない部分があることを述べている回答が見られた。次に示す自由記述のうち、2 つ目のコメントは納得できていないことをよく表している。

- ・自分が元々思っていた使い方ではない、正しい使い方で添削されていると思った。
- ・He や She であらわせないのに It にしているのだがそれを彼や彼女で表しているのは個人的には納得いかない部分であった。

学生たちは、DeepL Write の修正結果を見て、自身の英語力の課題を認識しやすくなっている。特に「文法の弱点」、「語彙力の不足」、「表現の洗練」といった点が浮き彫りになった。加えて、修正によって文の流れが良くなることを実感し、より伝わりやすい文章を書けるようになりたいという意識が高まっていた。

設問 7 は授業での DeepL Write の利用についてで、図 7 のような結果が得られた。

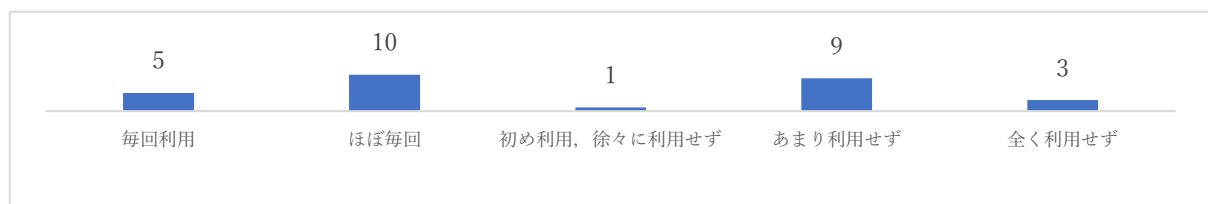


図 7. 授業での DeepL Write の利用について(n=28)

「毎回利用した」と「ほぼ毎回利用した」が合わせて 15 人で約半数以上が利用していた一方で、「あまり利用せず」と「全く利用せず」が合わせて 12 人で半分弱の学生が利用していないことが分かる。1 人が徐々に利用していないというのも興味深い。

設問 8 は設問 7 で「利用した」と回答した学生が DeepL Write を利用した感想を自由記述で回答するものであった。回答から次の 4 つのカテゴリーに分類することができた。

(1) 便利さと活用意欲

多くの学生が DeepL Write を「便利」と評価し、今後も使用したいと考えていることが分かった。特に、修正が素早く分かりやすい点が好評のようである。次の 2 つの回答はいずれも最後に「便利」が見られるが、後者はより具体的な回答となっている。

- ・自分が書きたいと思った文を英文にすぐに直してくれるので便利。
- ・アプリでカンミスや冠詞の間違いなどすごく分かりやすく訂正できるので便利。

(2) 学習支援の実感

英語学習において文法や単語の知識を再確認できる点や英語表現の仕方を学べる点の評価する意見が多く、学習ツールとして有効と感じた学生が多いことが分かる。次の回答例はいずれも「思い出す」がキーワードとなっている。

- ・忘れていた英文法のルールや英熟語が思い出されたため、知識が深まった。
- ・分からない言いまわしを提示してくれる文で見直すことにより忘れていた使い方を思い出すことができた。

(3) AI 技術への感想と懸念

AI 技術の進歩に対する驚きや称賛の一方で、人間の役割が減ることへの懸念を示している意見も見られた。便利さと学習の補助機能を評価しつつも、依存しすぎることへの不安を感じる学生もいることが分かる。次の回答はまさにこのことを示している。

- ・AI ってやはり素晴らしいものだと感じるとともに、これを使えば簡単に文章が書けてしまい、どんどん人の役割を奪っていくのではないかと心配に感じる。

(4) 難しさや使い勝手の問題

一部の学生は難しさや他のツールとの比較に言及しており、DeepL Write の修正が理解

しにくい、または自分の英語力の不足を痛感したという意見が見られた。簡潔な回答であるが、2つの回答を示す。

- ・難しかった。
- ・他のアプリの方が楽だなと感じた。

DeepL Write は、多くの学生にとって英語学習をサポートする便利なツールとして評価されている一方、使いこなし方や学習への影響について、考え方に差があることが分かる。

設問 9 は設問 7 で何らかの形で DeepL Write を利用したと回答した学生が引き続き利用したいかたずねるもので、図 8 はその結果を表したものである。

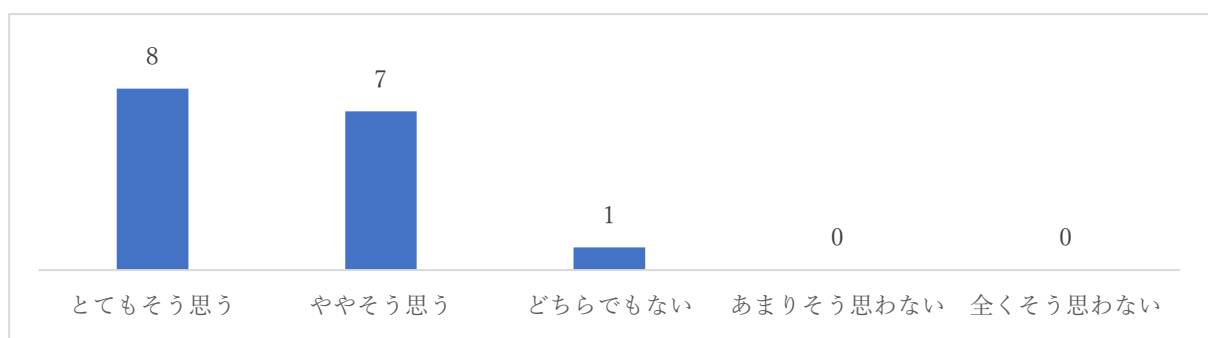


図 8. DeepL Write の今後の利用について($n=15$)

図 8 で示したように、「とてもそう思う」が 8 人、「ややそう思う」が 7 人、「どちらでもない」が 1 人で、否定的な回答は見られなかった。ほとんどの学生が、引き続き利用する意思を持っていることが分かる。

設問 10 は設問 9 で選択した回答について自由記述で回答するもので、次の 5 つのカテゴリに分類することができた。

(1) 便利さ・正確さを評価する意見

DeepL Write の便利さと、文法や表現の正確さを評価する意見が多い。特に、自分が考えた英文よりも正確なものが作れることをメリットと感じている。次の 2 つの回答にはいずれも「正確だ」という表現が見られる。

- ・ 便利だと思ったため。自分の考えた文法より正確だと思ったため。
- ・ DeepL Write の方が正確だと思ったから。

(2) 英作文の学習ツールとしての有用性

英作文の質を向上させるツールとして有用だと感じる学生が多い。特に、ニュアンスの違いに気づける点や、スペルミスの防止などが評価されている。次の2つの回答のうち、1つ目はニュアンスの違いの気づき、2つ目は有用性を示している。

- ・ 英作文を書くうえで細かな文のニュアンスの違いに気づくことができるから。
- ・ より自分の書く力を身につけるには使いやすく有効的だと思ったから。

(3) 自己学習・自己確認のための活用

自分の英語力を確認・向上させるためのツールとして活用したいという意見や完全に頼るのではなく、補助的に使いたいと考える学生もいる。

- ・ 今後も英語を学ぶ機会があると思ったから、自分が考えた英語の文章は正しい文法で書けているのかを確認したいから。
- ・ 分からない単語を調べられない。添削としては引き続き利用したい。

(4) 言い換え・表現のバリエーションの発見

英語の言い換えやバリエーションを学べる点を評価する意見、単に正しい文法を知るだけでなく、表現の幅を広げるのに役立つと感じている。次が具体的な回答である。

- ・ 同じ意味で異なる単語・表現をだしてくれたから。

(5) 言いたいことを素早く表現できる

英語表現に困ったときのサポートツールとして活用したいと考えている。英語以外の言語にも対応している点を評価する意見もある。

- ・ なんて英語で言えばいいのか分からないときに使いたいと思ったから。
- ・ すぐに言いたいことを英語に限らず他言語でも翻訳してくるから。

この結果から、DeepL Write は英語学習のサポートツールとして幅広い用途で活用されていることが分かる。

問 11 は設問 7 で DeepL Write を利用しなかった理由を自由記述で回答するもので、次

の4つのカテゴリーに分類することができた。

(1) 自力で書くことの重視

学習効果を高めるために、あえて AI ツールを使わなかったという学生が多い。間違えることで学ぶ姿勢を重視し、「自分の力でどこまでできるか試したい」と考える傾向が見られる。次の2つの回答に自力で書くことを重視していることが伝わってくる。

- ・自分の力で書いてみたかった。
- ・自分の力でやって間違っていた部分や勘違いしてた部分を見直したかったから。

(2) 他のツールの利用

DeepL Write の代わりに Google 翻訳を使ったという意見も多い。Google 翻訳の方が見やすい・使いやすいと感じる学生もいることが分かる。具体的な回答例を示したい。

- ・Google 翻訳で単語だけを調べて書いたから。
- ・Google 翻訳が一番見やすかったり、使い方がよく分からないため使用しなかった。

(3) 時間や環境的な問題

時間がなかったり、インターネット接続の問題、データ通信量や充電を気にする学生もいた。特に環境要因による未使用のケースも一定数あることが判明した。次の2つの回答のうち、1つ目は「時間」、2つ目は「接続の問題」を示している。

- ・文を考えることに必死で、利用する時間がなかった。
- ・インターネットの接続が悪かったので利用できなかった。調べるより、まずは自分で全部書いてみたかった。

ここまでの回答から自分の実力を理解するため敢えて利用しなかったり、使い慣れている Google 翻訳の方が使い勝手が良かったことが窺える。

最後の設問は授業で AI ツールを利用して作文を行うことについて自由に記述するもので、4つのカテゴリーに分類することができた。

(1) AI ツールの利用に肯定的（積極的活用）

AI ツールを積極的に授業で取り入れるべきと考える学生が多数であった。「便利」、「効率的」、「学習効果がある」といった理由が主に挙げられている。具体例を3つほど示し

たい。

- ・ 便利なものはどんどん使った方が良いと思うので、授業で AI ツールを用いることはとても良いと思います。
- ・ より正しい文法で相手に伝わりやすい文章を書けると思うので AI ツールを利用するのはアリだと思います。
- ・ 自分が分からない点、不安な点をすぐに添削してくれるので、積極的に利用推奨すべき。

(2) AI ツールの補助的活用

AI に頼りすぎるのではなく、自分で考えた上で活用するのが望ましいという意見も多い。答えをそのまま受け取るのではなく、自己学習の補助として使うべきという考え方が見られる。

- ・ 全くまっさらなものを翻訳するのは意味がないけれど、自分が考えたものを直してもらったり、分からない表現を調べることに使うのはとても有意義だと考えます。
- ・ 自分で考えた後に答え合わせのように使用するのが良いのではないかと考えた。

(3) AI ツールの活用に慎重

AI ツールを使用する前にまずは自分で考えることの大切さを指摘し、活用に関しては慎重に行った方が良いという意見が見られた。AI ツールよりも教員を信頼し、添削してほしいという意見もある。

- ・ あくまで AI ツールは分からなかった時の最終手段として利用するのはいいと思う。少なからず、自分で考えなければ意味がないので頻繁に利用はしない方がいい。
- ・ AI ツールが完璧に正しいと思えず、どうしても先生に訂添（原文のまま）してほしいと思ってしまう。

(4) AI ツールを授業に取り入れることに対する興味・好奇心

「最新技術を学べる」「新しい学習方法として面白い」と感じる学生もいる。学習以外

の観点で AI ツールの活用に興味を持っている様子うかがえる。

- ・最新技術を取り込んだ授業という風にとらえているので、時代の最先端にいるような気分がしてとてもおもしろく思う。
- ・限られた時間の中で書く英作文は単語が出てこなかったり、構成を練るのが難しいので AI ツールがあると突破口になってスラスラ書ける印象があった。

全体としては、AI ツールの活用に肯定的な意見が多く見られたが、使い方には工夫が必要で、一部で活用に関して懸念を抱く学生も見られた。また、学習意欲を向上せる役割としても機能している。授業で AI ツールを活用する際には、「使い方の工夫」と「適度なバランス」が重要であることが分かる。

7. 考察

本研究では 9 つの研究課題を設定した。研究課題 1「学習者は英語学習で AI ツールを使用したことがあるか。あればどのようなツールを使用しているか。」については、本実践以前に 85%以上が使用したと回答し、ほとんどの学生が使用している実態が見られる。またツールについては Google 翻訳を最も多く使用していた。Google 検索が広く普及していることから、学生の間では Google 翻訳を手軽に使用していることが予想される。

研究課題 2「学習者は授業受講以前に DeepL Write を知っていたか。」については 4 分の 3 の学生が DeepL Write については未知であった。そんな状況の中で学習者は DeepL Write を使用し、訂正フィードバックを受けたことになる。後者に関連する研究課題 3「学習者は DeepL Write の修正案についてどのように感じているか。」については「文法力の欠如」や「語彙力不足」を認識している学習者も見られるが、書いた英文が洗練されている印象を持っている学習者も多く見られた。中には修正結果に対する驚きや疑問を示す学習者も見られた。

研究課題 4「学習者は授業内でどの程度 DeepL Write を利用していたか。」については「毎回利用した」と「ほぼ毎回利用した」が合わせて約半数以上で、「あまり利用せず」と「全く利用せず」が合わせて半分弱であり、利用者の方が上回った。利用していた学

習者の意識に関する研究課題 5「授業内で DeepL Write を利用した学習者は、どのような感想を持っているか。また、継続して利用する意思があるか。」については言いたい表現が素早く表現でるといった利点や DeepL Write の文法の正確さを指摘する声があがっていた。そして今後の利用についてはほとんどが肯定的な回答をしており、今後も積極的に利用することが予想される。一方、利用しなかった学生の意識に関する研究課題 6「授業内で DeepL Write を利用しなかった学習者はどのような理由で利用しなかったのか。」については、訂正フィードバックを通じて自分の実力を理解できるため、自力を重視して敢えて利用しないという学生が多く、また使い慣れている Google 翻訳の方が便利であると感じている学生も一定数存在することが判明した。

研究課題 7「授業で AI ツールを利用して英作文を行うことについてどう思っているか。」については、AI ツールの利用に対して肯定的な意見が多く見られたが、一部で慎重な意見も見られた。AI ツールが示す修正案はすべてが正確というわけではないので、使い方について指導者が適切な助言を与える必要もあるだろう。

研究課題 8「学習者は授業で取り組んだ英作文について楽しいと感じたか。」については、肯定的回答は約 60%で、否定的回答は 25%で、6 割以上の学生は楽しんで取り組んでいたことが分かる。ただし、この結果は DeepL Write を使えたことと関係があるかどうかまでは不明である。2 週間に 1 度英作文を書く必要があったのでライティングに苦手意識を持つ学生が否定的な回答をした可能性がある。

研究課題 9「学習者は授業を通じて書く力がついたと感じたか。」については、肯定的回答は約 64%で、否定的回答は約 21%で、6 割以上の学生が書く力がついたと実感していることが分かる。利用していた学生の中には DeepL Write の支援を受けることができたこともあり、そう感じているのかもしれない。

本来ならば DeepL Write が示した修正案について学生に再度検討させた上で、書き直した原稿を再提出させたほうが良いと思われるが、今回は実施できなかった。

8. おわりに

本研究では、AI 駆動型の文章支援ツールの DeepL Write を学習者に利用することを可能とした環境下で英作文を行った。学生が書いた作文を筆者が読み上げ齟齬がないか確認の上、DeepL Write に添削させ、それをプリントアウトし、DeepL Write 未修正部分を修正し、コメントまで添えるという訂正フィードバックであった。教員の負担という観点では効率の悪い訂正フィードバックだったかもしれない。しかし授業やアンケート結果から学習者が満足していることが感じられた。最初から電子データで提出させることで、効率化が図れるので今後取り組んでみたい。

教室内で支援ツール等が得られない環境下でも次のような方法が考えられる。まず、学習者に英作文を書かせ、提出させる。教員はその英作文を読み、修正すべき箇所に下線を施すなどして、コメントを添えて返却する。学習者は下線やコメントを参考に自宅等で AI 支援ツールを用いて書き直し、再提出する。教員は書き直された英作文を読んでコメントし、学習者に返却する。この過程で、教員が添削する負担は減るはずだ。学習者は見直し作業を通じて、語彙や語法の正しい使い方を認識し、英語力の向上につながる可能性がある。

謝辞

本稿の完成にあたり、貴重なご意見をくださった匿名査読者、並びに論文審査委員会の方々に、この場をお借りして厚く御礼申し上げます。

参考文献

- 甲斐順(2024). 「ライティングツールを利用した英作文添削比較」『言語学習と教育言語学：2023 年度版』 9-20.
- 佐野富士子(2013). 「第 19 章 ライティング」 JACET (大学英語教育学会) SLA 研究会編『第二言語習得と英語科教育法』 (pp. 248-261). 開拓社
- 佐野富士子(2023). 「第二言語習得研究の発達の概観と現代的課題—EFL 学習者の英語力育成を目指す教室での第二言語習得—」『JACET 関東支部紀要』 10, 5-26.

- 鈴木渉(2017).『実践例で学ぶ第二言語習得研究に基づく英語指導』大修館書店
- 新美德康・梅木璃子(2024)「自動ライティング評価ツール Grammarly を援用した英語ライティング指導の探究的実践」『リメディアル教育研究』18, 69-80.
- 山下美朋・山中司・杉山滉平(2024).「AI ツールを活かした英語ライティング授業: 英文作成支援ツール Transable を導入して」『立命館高等教育研究』24, 75-87.
- DeepL.(2024). 「DeepL について」
<https://support.deepl.com/hc/ja/articles/6318834492700-DeepL-Write%E3%81%AB%E3%81%A4%E3%81%84%E3%81%A6>(2024 年 11 月 7 日閲覧)
- Ferris, D. (1999). The case for grammar correction in L2 writing classes: A response to Truscott (1996). *Journal of second language writing*, 8(1), 1-11.
- Gold, J. (2023). Fostering self-directed writers using AWE and WBMT: A pilot case study. *Journal of the school of languages and communication, Kobe University*, 19, 1-27.
- Hyland, F., Nicolás-Conesa, F., & Cerezo, L. (2018). Key issues of debate about feedback on writing. In R. M. Manchón, & P. K. Matsuda (Eds.), *Handbook of second and foreign language writing* (pp. 433-452), De Gruyter Mouton.
- Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language learning*, 46(2), 327-369.
- Truscott, J. (1999). The case for “The case against grammar correction in L2 writing classes”: A response to Ferris. *Journal of second language writing*, 8(2), 111-122.

大学英語教育における ChatGPT の導入 :

ライティング学習支援効果と課題の検証

Implementation of ChatGPT in University English Education: Its Effects and Challenges in Writing Skill Development

古藪麻里子 (東洋大学)

藤城晴佳 (東洋大学)

町村貴子 (東洋大学)

キーワード : ChatGPT、生成系 AI、ライティング指導

Abstract

This study aims to introduce generative AI into university-level English writing classes and delineate its effects and challenges. In practice, four class sessions utilizing ChatGPT as writing advisors were conducted for approximately 300 first-year university students. The results indicate that ChatGPT provided flexible advice, improved the quality of students' writing, and promoted self-correction abilities. However, it was also confirmed that lower-level students needed support to understand and reflect on the AI's suggestions fully. Implications for educational practice suggest that generative AI can improve writing students' skills when carried out under appropriate instructional design. As for prospects, effective approaches such as continuous practice with short writing tasks should be explored and implemented. Additionally, verifying long-term effects and adaptability to other subjects and elucidating students' cognitive processes could enhance the practical application of generative AI in education.

1. はじめに

1.1 研究の背景

2022 年 11 月にリリースされた ChatGPT が、人工知能の新たな可能性を世界に示して以降、Bing Copilot や Bard といった新たなサービスが次々と現れ、生成系 AI への注目が急速に高まっている。ただし、これらの進化はまだ始まったばかりであり、プライバシー保護などの理由から一定の制限が設けられる可能性があるものの、今後も生成系 AI の進歩は止まることなく続いていくことは確実である。

これを受け、教育分野における生成系 AI 技術を教育に統合することに積極的な立場を支持する論調が高まる一方で、教育界全体がこの技術革新のペースに適応していると言い難い実情が観察される。学生側も、ChatGPT をはじめとする生成系 AI の利用経験は一般的であるが、このような技術を自らの学習にどう活かすべきかについて、明確な理解が得られていない状況である。また、生成系 AI の不適切な使用は、自主的な学習プロセスに悪影響を及ぼす可能性があるにもかかわらず、この技術の誤用に対する十分な認識が欠如していることが懸念されている。

急速に普及する生成系 AI であるが、教育分野におけるその応用に関する取り組みの具体例はまだ少ない。次節では、既存の先行研究を概観し、本実践の成果がどのように位置づけられるかを明らかにしていく。

1.2 先行研究

外国語教育における生成系 AI の影響について、金丸（2023）は AI を学生が主体的に活用しながら学習を進めることの重要性を述べている。学生は AI の出力を参考にしつつも、自らの理解を深め、基礎知識や技能を身につけていく必要があるとしており、教員の役割は、学生が外国語の文法、語彙、構造などの基盤を確実に習得できるよう、適切な指導を行うことにあるとしている。

小論文のライティングにおける ChatGPT の具体的な活用方法とその効果については、Shields（2023）の研究が挙げられる。Shields は、ChatGPT が生徒の考えを整理する際の支援ツールとして有用であり、特に小論文の構造を提示し、文法、スペリング、語彙のフィードバックを行うことで、生徒が自分の言葉で作文する際の一助となり得ると述べ

ている。これにより、生成系 AI が作文の品質向上に寄与する可能性が示唆されている。

さらに、ライティングにおいて、生成系 AI と人間の得意分野を適切に識別し、相互に補完しながら作業を進めることの重要性が柳瀬（2023）によって提示されている。柳瀬は、日本語話者による短期間の英語論文執筆プロセスとして、i) ストーリーの戦略的設計、ii) 日本語での執筆後の AI による英語翻訳、iii) AI への客観的な英文体改善指示、iv) 著者による最終稿の確認、という一連の手順を提案している。これらの流れを補完する形で、岩中（2023）は、AI が言語形式のフィードバックを、教員がエッセイの内容のフィードバックを担当することで、相補的な役割を果たし、効率的で高い教育成果が得られる可能性を示唆している。

以上のように、これまでの先行研究では、学生が主体的に使用することの重要性が共通して指摘されているものの、AI にどこまで指導やフィードバックを委ねるのかについては、まだ一致した見解は得られていない。

一方で、学生側は AI を使用することについてどう考えているのか、また生成系 AI の活用の現状について、古藪ほか（2024a）の報告がある。ここでは、データサイエンス系学部の英語必修科目でライティング課題に取り組む学生の約半数が ChatGPT を利用していることが確認されている。また、使用しなかった学生グループにおいても、多数が将来的な使用に対して肯定的な意向を示しており、学生間での生成系 AI を使った学習活動への関心が高いことが示唆された。さらに、収集されたデータからは、ChatGPT の適切な使用方法について授業内での指導を望む学生が全体の75%以上に上り、学生たちの中で生成型 AI を効果的に活用したいという強いニーズが示唆されている。

以上の先行研究を踏まえると、生成系 AI を教育への補強として適切かつ効果的に活用するためには、AI を教育内容に組み入れた教材の開発とその運用および分析が急務であると言える。本研究では、このような背景をもとに、ChatGPT を活用した英語ライティング授業の実践報告を行うことを目指す。

1.3 本研究の目的

英語のライティングの授業で ChatGPT を活用したカリキュラムを実践し、その学習支援効果と運用上の課題について検証することが、本研究の目的である。

2. 授業実践

2.1 実践概要

実践期間は 2023 年度の 9 月から 1 月までで、東京都内の私立大学のデータサイエンス系学部に所属する 1 年生 303 名（8 クラス）が対象となった。対象科目はリーディング・ライティング演習Ⅱという週 1 回開講される英語必修科目である。なお、本研究は実践した大学の倫理審査委員会の承認を得て行われたものである。

2.2 学習環境

本学部では、全科目において、BYOD (Bring Your Own Device) を採用しており、各学生が自分のパソコンを持参して授業を受ける形をとっている。また 2023 年度より、学部教育全般に生成系 AI の活用を推進しており、学生たちは最新技術を活用し、効率的な学習を進められる体制が構築されている。例えば、学生はウェブ版の ChatGPT を直接使用するのではなく、Slack というコミュニケーションプラットフォームを介して有料版の ChatGPT 最新モデル(本研究においては ChatGPT-3.5 Turbo および ChatGPT-4 を使用)にアクセスが可能となっている。新入生には、入学ガイダンスや学期開始直後の情報関連必修科目で、Slack や ChatGPT の基本操作を学ぶ機会が提供されていたため、英語の授業でも比較的スムーズにこの技術を活用できる状況にあった。

2.3 カリキュラムおよび使用教材

2.3.1 カリキュラム構築

今回、ChatGPT をカリキュラムに導入するにあたって、AI を独立した単元として学習するのではなく、従来より行っていたライティング活動に ChatGPT を新たなサポートツールとして使用する、いわば統合型のアプローチを採用した。本カリキュラムにおいて

ChatGPT は、学生がライティング力や思考力を発展させるための文書作成のアドバイザーやライティングのトレーニングパートナーとして位置づけられている。今回の実践の目的は、AI の使い方を学ぶことではなく、AI をツールとして活用し、英語力や文章力を向上させることにある。したがって、ライティング指導のカリキュラムに合わせて、ChatGPT を使った活動を組み込むことが最も効果的であると判断した。

そのために、ライティングにおいて ChatGPT のサポートが有効な 5 つの活用法を特定し、それらがエッセイ作成の各ステップに組み込まれた授業を実施した（図 1）。

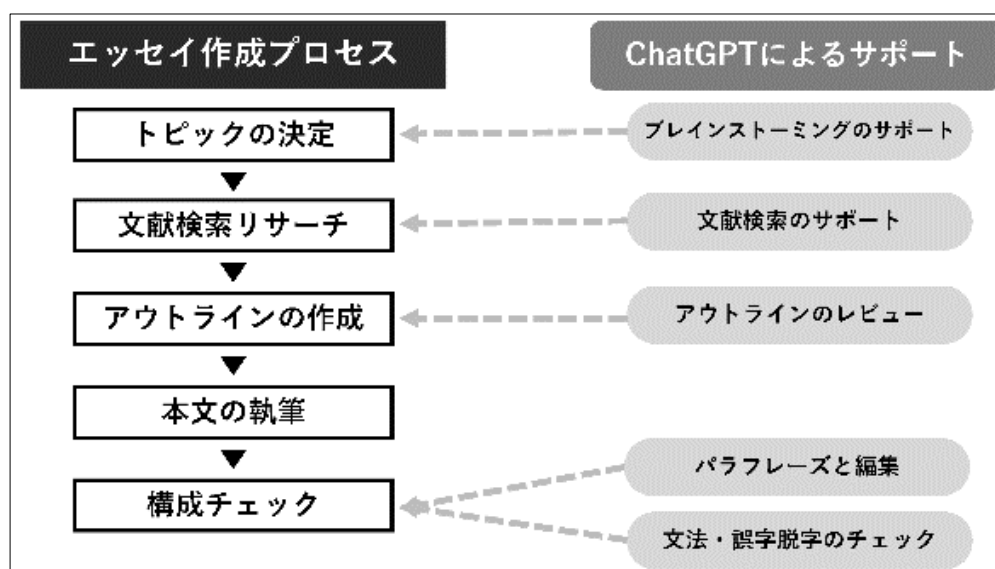


図 1. ChatGPT 5 つの活用法. 古藪ほか（2024b）より転載

実践は学期を通して 4 回の授業セッションを通じて展開され、指導には各回 20 分から 30 分が割り当てられた。

生成系 AI の使用に対しては「学生が自分で考えなくなる」という懸念が指摘され、教員や教育機関が教育への導入をためらう理由のひとつとなっているが、今回の実践が行われた学部では、生成系 AI を使うことによってむしろ学生の考える力を伸ばすことを目指している。このため、ライティング課題では「問題文を打ち込み、出力をそのまま貼り付ける」という行為を防ぐことを目的として、エッセイの計画段階で自らが考えて書いたアウトラインを ChatGPT にレビューさせ、その結果をエッセイに反映させること

を義務付けている。また、提出前に文法・誤字脱字のチェックも実施するように推奨したが、それ以外の項目に関しては、説明は行ったものの ChatGPT の使用は任意とした。

2.3.2 使用教材

実践に際し、既存のワークシートやパワーポイントスライドといったオリジナル教材に加え、新たに作成・導入した教材に『ChatGPT 活用ガイド』（A4用紙5枚）がある。このガイドには、ChatGPT の使用ルール（図2）、効果的な活用方法、注意点、そして AI に送るプロンプト（指示文）の例（図3）が含まれている。プロンプトは対話の内容や質に大きく影響し、表現を少し変えるだけで意図しない回答が返ってくる可能性があるため、特に注意が必要である。学生が単に答えを得るのではなく、自ら考え、思考力を鍛えることを促すために、プロンプトの作成には慎重を期した。プロンプトの例は学生の理解を助けるため日本語で記述し、習熟度の高いクラスには英語でプロンプトを書くことも推奨した。ただし、ライティングスキルを高めるため、エッセイの初稿は、AI の翻訳機能を使わずに英語で執筆するよう指導している。

主な使用ルール

- ① 必ずログ（AIとのやりとりの記録）のスクリーンショットを貼付する
- ② AIの出力をただコピー＆ペーストするのではなく、内容をよく確認した上で、必要であれば自分で再入力すること
- ③ AIに過度な依存をしない
（AIの意見を自分の意見とする、情報を鵜呑みにするなど）
- ④ 検索ツールとして使用しないこと

図2. 主な使用ルール

e. 文法・誤字脱字のチェック

- 提出前に文法的な誤りやタイプミスなどがないかの最終確認（必ず修正した箇所を明示するような指示を出しましょう）
 - 「以下の〇〇に関するエッセイを読み、文法的な誤りや誤字脱字がないかを確認してください。間違っている点は箇条書きにし、修正案とともに提示してください。『ここに文章を挿入します』」


 **Point!** 修正後の文章だけでなく、どこを直すべきかを必ず説明してもらいます。指摘されたポイントをひとつひとつ確認しながら、手動で修正しましょう。AIの答えが必ずしも正しいとは限りませんので過信は禁物です。

図 3. プロンプト例

2.4 調査方法

本調査では、教員の視点や実際の使用記録（ログ）も含め多角的に分析することで、これからの授業実践やカリキュラム開発において、より具体的な提案ができると考え、以下2つの調査項目から分析を行った。

2.4.1 調査項目1：ライティングの質の分析

調査対象の学部では、入学前にすべての1年生に対してJIEM（株式会社教育測定研究所）が開発・運営を行う英語検定試験CASEC（Computerized Assessment System for English Communication）の受験を義務付けでおり、本調査の対象科目ではその得点をもとに習熟度別学習を実施している。

まず、ChatGPTの使用が学生のライティングに影響を与えているか検証するため、ライティング課題の評価点を前年度（ChatGPTの使用なし）と比較した。課題内容は、いくつかのトピックの中からひとつを選び、それに関して、250語以上で4パラグラフ構成のPros & Cons Essayを書くというものであった。調査対象となった年度（2022年度、2023年度）のライティング評価は、共通の採点基準に基づいて実施された。評価にはアウトラインが含まれ、チェックリストに基づき総合的に採点された。具体的な採点ポイントは、基礎的なライティングルール（例：パラグラフのインデントなどの4項目）、文献引用の正確さ（4項目）、そしてコンテンツの質（例：タイトルやトピックセンテンス

の適切さ、pros と cons の配分などの 6 項目) の計 14 項目であった。さらに、採点教員間で評価基準を共有し、事前に基準の詳細を確認するための打ち合わせを行うことで、評価の一貫性が保たれるよう配慮している。また、習熟度により評価点に違いがあるか調べるために、前述した CASEC のスコアをもとに、2022 年度課題提出者 249 名、2023 年度課題提出者 284 名をそれぞれ、高得点群 (上位 25%)、中得点群 (中間 50%)、低得点群 (下位 25%) の 3 群に分けて分析を行った。各群の CASEC 平均点と TOEIC 換算した詳細は以下の通りである (表 1)。

表 1 群別 CASEC スコア

年度	群	平均点	最高点	最低点
2022	高得点群	633	711 (840)	584 (520)
	中得点群	519	582 (515)	449 (350)
	低得点群	339	448 (350)	41 (15)
2023	高得点群	643	823 (905)	591 (530)
	中得点群	527	588 (525)	463 (365)
	低得点群	383	462 (365)	129 (80)

*括弧内は TOEIC 換算点

2.4.2 調査項目 2: 教員への聞き取り調査

学期終了後の 2024 年 3 月に授業を担当した教員 4 名へ聞き取り調査を行った。本実践は実施学部においても新しい実践であるため、教員を対象に聞き取り調査を行うことで、授業の雰囲気や教材のレベル、学生の反応や、運用上の問題点等についての内情を明らかにすることを目的とした。

聞き取りの内容としては、① ChatGPT を導入したことに対する考え、教えてみた感想、② 前年度との違い、③ 学生の理解度、成果物について、④ 来年度の実践に向けた問題点や改善方法である。

3. 結果と考察

3.1 授業運営の円滑性と運用上の課題

3.1.1 授業運営の円滑性

全4回の授業セッションを通じて、授業運営は概ね順調であった。技術的なトラブルとしては通信エラーにより一部で ChatGPT にアクセスできない状況が発生したが、その場合はウェブ版の ChatGPT に切り替えることで対応した。操作に手間取り授業が停滞することはなく、学生たちは抵抗なく利用していた。このスムーズな対応は、他科目での使用経験が大きく影響していると考えられる。もし ChatGPT に不慣れな学生が多かった場合、より丁寧な指導や教員による個別フォローアップが必要となる可能性がある。

一方、課題として浮上したのは学生によるログの申告である。ライティング課題において、すべてのログのスクリーンショットの添付を義務付けていたにもかかわらず、対話の一部しか報告していないなど添付が不完全であると思われるケースが複数のクラスで見受けられた。ログが完全でない場合、学生自身が考えた文章やアイディアの範囲を正確に判別することが困難になる。また、対話を重ねることで学生の思考を深めることが授業の目的の一つであるが、その対話の過程が見えないと、学生が課題を AI 任せにするリスクが生じる。このことから、完全なログの提出は極めて重要であるが、この意図が一部の学生には十分に伝わっていなかったようである。さらに、翻訳 AI を利用したと疑われるケースも見受けられ、将来的には翻訳 AI も含め、すべての AI とのやり取りを報告することを徹底する必要がある。

3.2 ChatGPT 導入前後でのライティングの質の変化

ChatGPT を活用することがライティングの成果に与える影響を検証するため、学期末に実施されたライティング課題の評価点を前年度と比較した。両年度の課題は、同一内容および同一採点基準に基づいて評価された。なお、この採点基準では、意味が通じる範囲でのマイナーな文法間違いやスペルミスは減点対象とされていない。課題は 15 点満点で採点され、全体の平均点は 11.06 点から 11.74 点に上昇した。 t 検定（片側）の結

果、この得点の上昇は統計的に有意であることが確認された ($p=0.031$)。さらに注目すべき点として、高得点群、中得点群、低得点群のすべてで平均点が上昇したことが挙げられる。これにより、ChatGPT を活用したライティング指導が、学生の英語力を問わず効果的であったことが示唆される。

表 2 ChatGPT 導入前と導入後の課題平均点の比較 (n=525)

2022 年度(n=249)		2023 年度(n=276)		t 値	p 値	F 値
平均	標準偏差	平均	標準偏差			
11.06	4.86	11.74	3.45	1.85	.031*	< .001

* $p < .05$

指導・採点にあたった教員の印象も、この統計結果を裏付けるものであった。ChatGPT による提出前のチェックを行ったことで、誤字脱字や文法的なミスが大幅に減少したことが確認された。ただし、それ以上に、文章の流れがスムーズになり、従来よりもライティングの品質が向上したと教員は感じていた。特に、課題とともに提出されたアウトラインが改善され、その結果、前年度よりも適切な構成で論理的に書かれた課題が増加したという点で、全教員が一致していた。つまり、ChatGPT をライティングツールとして活用することにより、英語の文章力そのものだけでなく、文章構成力も向上したと考えられる。

3.3 ログの分析結果

学生が ChatGPT とどのような対話を繰り返しているかを調査するために課題に添付されたログを分析した結果、いくつか興味深い点が浮かび上がった。以下では、具体的なログの実例を挙げながら、これらの点を紹介する。

まず、全体として大多数の学生が授業内で指示された ChatGPT の活用方法に従って問題なく使用していたことが分かった。これは、3.1.1 述べた他科目での使用経験に加え、『ChatGPT 活用ガイド』でプロンプト例を示していた効果が大きかったと考えられる。一方で、ガイドの例以外のプロンプトを使用した学生は、意図と異なる回答が出るケースもあり、効率的な使用のためには、学生に対して適切に吟味されたプロンプトを提示することが重要であると確認された。

なお今回の実践では、ChatGPT との対話に使用する言語の指定を行わなかった。大多数の学生は日本語で対話していたが、上位～中位クラスの中には英語でプロンプトを作成している学生もいた。その場合、【ログの記録①】(図4) 前半部分のようにプロンプトの英語表現に多少の不正確さがあっても、適切な内容の回答を得ていた。

ChatGPT のレスポンスはほとんどのケースで理にかなったものであり、文法や語用だけでなく、パラグラフの書き方をはじめとするライティングの基本的な考え方を含めて、教員から見て不適切と思われる回答は見受けられなかった。ChatGPT が行ったアドバイスの具体例については【ログの記録②】(図5) を参照されたい。

次に、学生たちはどのような場面で ChatGPT のサポートを得ていたのかを分析した。必須とした「アウトラインのレビュー」以外で一番多く使用されたのは「文法・誤字脱字のチェック」であったが、「ブレインストーミングのサポート」に使用する学生も多数見られ、知識量の少ない学生が与えられたトピックに関して何を書くべきか分からない場合に利用されていた。また【ログの記録①】(図4) の後半に見受けられるように、「どのように直せば1つ上のランクの文章を書けますか」とライティングのレベル向上を目指すアドバイスを求める学生もいた。これにより、ChatGPT が単なる「間違いのチェックツール」に留まらず、学生の自主性に基づく高次のライティングスキル向上に寄与する可能性が示された。

今回の実践結果では ChatGPT によるライティングサポートは、教員によるフィードバックや他のソフトウェアとの違いを特徴として見せた。第一に【ログの記録①②】(図4、5) に見られるように、文書構成のアドバイスやアイデア出しのサポートは、対話型

の AI であればこそ可能となった機能である。また【ログの記録③】(図 6) のように語用や表現のニュアンスに関する個々の事例に応じたアドバイスも提供され、既存の添削ソフトでは実現できない文法の正確性を超えた解説が提供された。さらに、個々の学生に対して執筆過程でこれだけの詳細な改善点を指摘することは、人間の教員には時間的・労力的に困難である。これらのことから、今回の実践を通じて、ChatGPT によるライティングサポートはその量・質において生成系 AI ならではのものであることが確認できた。


ChatGPT を学習に導入するにあたり、教員は「ChatGPT の指摘をそのまま受け入れるのではなく、自ら考え納得できるまで対話を続けること」を指導した。これは、学生が AI の出力を鵜呑みにするのではなく自ら深く考えることを奨励するためである。また、ライティング課題ではアウトラインのレビューを義務付けた。これは、アウトラインを添付させずに最終稿のみを提出させると、すべてを AI 任せにする学生が出るリスクがあるためであり、アウトラインの段階で AI からフィードバックを得ることで、論理的な問題や構造の欠陥に事前に気づき、再考することで回避できると期待されたからである。しかし、この措置だけでは ChatGPT の使用が学生の思考プロセスにどのように影響を与えたかを十分に把握することはできない。ChatGPT を活用することで学生が自分の考えを深めているかを検証するため、学生がどの程度の回数、どのような内容で ChatGPT との対話を重ねているかを調査した。結果として、1 ターン（学生の問いかけに対して ChatGPT が回答をする）のみでやり取りが完了するケースは少なく、むしろ多数のターンを重ねる学生が多く見られた。中には、30 ターン以上対話を続けた例や、友人に話しかけるような口調で様々なアイデアを挙げながら、どのように書くべきかを相談する様子なども見られた。また、【ログの記録③】(図 6) のように、ChatGPT のアドバイス内容に納得がいかなかった点を問い直し、その結果を最終稿に反映させているケースも確認できた。これらの例は、ChatGPT との対話を通じて、学生が批判的思考を発揮し、適切な英語表現を模索したり熟慮を重ねた文章を書いたりする能力が促進された過程の表れであると考えられる。このように、ChatGPT をうまく活用できている学生は、英語

力の高い学生だけでなく、中位から下位の学生にも存在しており、このことが高得点群、中得点群、低得点群のすべてで平均点が上昇した要因となったと推察される。

有効な対話を通してライティングの質を高めている学生がいる一方で、ChatGPT による修正案を無視する学生も多数見受けられた。その場合、例えば、「論拠を示すべき」「具体例を挙げるべき」といった明白に正当なアドバイスでさえも実行されていなかった。このことから、学生が自らの判断で「AI の示した方向性に従わない方が良い」と考えたのではなく、ChatGPT のアドバイスを正確に理解していなかった、あるいはその手間を回避したために修正を行わなかったと推察できる。このような学生は、低得点群だけでなくすべてのレベルに見受けられた。このことは、英語力と AI をライティングに活用する能力が単純に比例するわけではないことを示唆している。したがって、今後は ChatGPT の使用方法だけでなく、提案内容の理解とライティングへの反映をどのように支援するかを検討する必要がある。

Aさん
CASECスコア: 488, TOEIC換算: 390

What do you think of Drone deliveries? I want you to know pros and cons about it. Please write down within 6 item

 Drone deliveries hold promise for revolutionizing the logistics industry by offering faster, more efficient delivery services. Here are the pros and cons of drone deliveries within six items:
Pros (略)
Cons (略)

----- 中略 -----

添削前と比べてどのように直せば1つ上のランクの文章を書けますか？


 あなたのドローン配送に関するアウトラインをさらに向上させるためのアドバイスを以下に記します。これらのポイントは、より洗練された、上位ランクの文書を作成するのに役立ちます。
1. 具体的な例とデータの利用 (略)
2. 議論の深掘り: (略)
3. 比較と対比: (略)
4. 結論の強化 (略)
5. 視覚的要素の追加 (略)

図 4. ログの記録①

Bさん

CASECスコア: 463, TOEIC換算: 365

顔認証についてのpros&consエッセイを作成中です。アウトラインをレビューして、アドバイスをしてください。以下がアウトラインです
【アウトライン(略)】

日本語でおねがいします。



顔認証システムの普及について部分的に賛成している点を述べるというアウトラインは、よく考えられておりバランスが取れているように思われます。ただし、少し詳細な指摘をさせていただければ、以下のようなアドバイスがあります。

1. まず序論で、顔認証システムの基本的な概要と、それが広く普及している理由、または普及している背景について触れると良いでしょう。
2. ポジティブな点(a)
 - i. 伝染病防止の非接触・非対面というメリットを挙げていますが、より具体的な例や、それがどのように人々の安全や利便性に貢献しているかについての説明を加えると良いでしょう。
 - ii. 心理的な負担が少ないと言う点は非常に良い観察ですが、ここにも「人間が本能的に他人の顔を識別することに慣れている」という心理学的な側面を加えて補強すると、説得力が増します。
3. ネガティブな点(b)
 - i. 認証精度については、具体的な数字や研究結果を引用して、どれくらいの頻度でエラーが発生する可能性があるのか示すと、読者の理解が深まるでしょう。
 - ii. プライバシーについては、顔認証システム使われる際のデータ保護法や規制に触れることで、そのネガティブな影響をより具体的に示すはどうでしょうか。また、データ侵害のリスクや悪用の危険性についても触れると、懸念点がより明確になるでしょう。
4. 論点がバランス良く取り扱われていることを確認し、それぞれの項目に対して補強する具体例やエビデンスを用意することが重要です。
5. 結論では、これらの利点と欠点を総合して、顔認証システムの普及についてどのように感じるかをまとめ、将来的にどのような改善が望まれるかについて提案すると、より効果的なエッセイになるでしょう。良いエッセイ作成の参考になると幸いです。

図 5. ログの記録②

Cさん

CASECスコア: 597, TOEIC換算: 540

Self-drivingについてについてエッセイを書き、修正しました。何かスペルミスや文法などの間違い、さらに修正したほうが良い点があれば、教えてください。アドバイスは日本語をお願いします。以下がエッセイです。(以下は該当箇所のみ記載)

Leaving everything to the AI could lead to intentional accidents if the system malfunctions or is hacked.



"Leaving everything to the AI could lead to intentional accidents"では"intentional" (意図的な)ではなく"unintentional" (非意図的な)が正しい表現です。

intentional accidents の意図的な事故というのは、ハッカーが車をハッキングし、意図的に事故を起こさせるという意味合いだったのですが、よくないでしょうか？



あなたのエッセイの文脈において「intentional accidents」という用語は、ハッカーが車をハッキングし、意図的に事故を引き起こすというシナリオを指すために使われています。この場合、事故は故意にプログラムされた結果であり、表現としては適切です。しかし、この用語は一見すると矛盾しているように見えるかもしれません。通常、「事故」は不意やミスによるものと解釈され、故意に起こされるものではありません。そのため、もし読者に対してこの点を明確にしたい場合、エッセイに更なる説明を加えるか、「intentional accidents」の代わりに「hacking-induced accidents」や「malicious sabotage」といった表現を使用することで、意味がより明確になり、誤解を避けることができます。

学生の提出した文：

修正前：Leaving everything to the AI **could lead to intentional accidents** if the system malfunctions or is hacked.

修正後：Leaving everything to the AI **could lead to hacking-induced accidents** if the system malfunctions or is hacked.

図 6. ログの記録③

4. 結論

4.1 ライティング支援ツールとしての ChatGPT

今回の実践を通じて、ChatGPT が大学英語教育におけるライティング支援ツールとして有効であることが確認できた。ChatGPT の回答はほとんどのケースで理にかなっており、既存の添削ソフトウェアにはない柔軟なアドバイスを提供していた。教員がエッセイの執筆段階で詳細なアドバイスを提供することは時間的・労力的に困難であり、ChatGPT を用いることで学生は納得できるまで何度でも対話を繰り返すことができる点は、生成系 AI を用いたライティング教育の大きな利点の一つといえる。

学生は授業で扱った使い方を超えて ChatGPT を多様に活用し、独自の方法を模索していた。また、多くの学生が多数の会話ターンを重ねており、学生が AI の出力を無批判に受け入れるのではなく、自ら考える過程が促進された様子もうかがえた。その結果、提出された課題は誤字脱字や文法上のミスが少なく、エッセイの構造も論理的なものが多かった。全体としても前年度と比較してクオリティの高いものが多く、課題の評価点の平均がすべての習熟度で前年度を上回っていた。教員の印象も同様であり、今回の ChatGPT の導入がライティングのクオリティの改善に役立ったと考える。一方、特に下位クラスの学生には ChatGPT の提案内容の理解とその反映に課題があり、さらなる支援が必要であることも確認された。

4.2 今後の課題

今回の実践により、ChatGPT をはじめとする生成系 AI を用いた授業は、適切な授業設計（例：ステップごとに使用方法を解説し、プロンプト例を提示する）のもとで行えば、ライティングの質の向上や学生の自律的な思考力の促進に寄与することが明らかとなった。今後は、より効果的な ChatGPT の活用方法を探るとともに、課題点を克服するための実践や効果の検証が望まれる。

まず、本実施では、学生の英語習熟度と ChatGPT との対話内容との関連性に関する詳細な分析は行っていない。しかし、観察の結果、中位から下位の英語力の学生でも、ChatGPT を効果的に使ってライティングを向上させられることが分かった。一方で、すべての習熟度レベルにおいてフィードバックを十分に活用できない学生の存在も確認さ

れた。これは英語の理解力以外の要因がフィードバックの適切な反映を妨げている可能性を示唆しており、これらの要因の追求や、習熟度と AI 対話の関係性について、さらなる詳細な検証が求められる。

今後取り組むべき実践的課題には、ChatGPT の提案を理解し反映することが難しい学生への支援が挙げられる。具体策として、教員からのフィードバックをステップごとに細かく段階的に提供する方法が有望である。また、「自分で第 1 稿を考える」→「AI のレビューを受ける」→「修正する」というプロセスの習慣化を、短いライティング課題を通じて頻繁に実施することが望ましい。

また教材に関しては、ChatGPT からより質の高いフィードバックを得るため、学生に提示する最適なプロンプトを継続的に模索することが重要である。今回の実践結果から、ChatGPT がライティングのトレーニングパートナーとして有能であることが確認できたため、今後さらに使用を重ねることにより、学生の英語力やライティング力の向上が期待される。

一方、生成系 AI を用いた外国語教育の効果に関する検証は依然として不足しているため、今後の研究課題とする必要がある。検証すべき点としては、中長期的にどのような学習効果がもたらされるか、課題における得点だけでなく、学生の文法の正確性およびライティングスキル自体の向上が見られるかどうか、さらには英語以外の科目における適用可能性が挙げられる。また、利用者の認知プロセスの解明も必須の課題である。このプロセスを研究するための最適な手法は今後の研究によって決定されるべきであるが、ログデータを対象とした分析は有力な手法のひとつであろう。加えて、利用者へのアンケートやインタビューなど多様な方法を用いた総合的な検証が望ましいと考える。

今後、ChatGPT をはじめとする生成系 AI の活用は教育現場においてますます拡大していくものと予想される。教育者は、絶え間なく進化する技術に適応しつつ、最先端の動向を継続的に追い求める姿勢が求められる。今後の研究成果が大いに期待される一方で、実践と研究を通じて得られる知見を蓄積し、それらをもとに技術を効果的に活用する取り組みを一層強化していくことが肝要である。

参考文献

- 岩中貴裕 (2023).「英語教育における人口知能(AI)の活用—その可能性と大学における英語教育の将来についての考察—」『山口県立大学学術情報』 16, 109-119.
- 金丸敏幸 (2023). 「言語生成 AI は外国語教育に何をもたらすか?」『大学教育と情報』 2023(2), 16-17.
- 古藪麻里子・藤城晴佳・町村貴子 (2024a).「英語ライティング課題における学習者の ChatGPT 利用状況調査」『JACET 授業学ジャーナル』 4, 1-21.
- 古藪麻里子・町村貴子・藤城晴佳 (2024b, September). 「生成系 AI で変わるライティング指導—ChatGPT を活用した授業実践」『英語教育』 73(8), 36-37.
- 柳瀬陽介 (2023).「AI を活用して英語論文を作成する日本語話者にとっての課題とその対策」『情報の科学と技』 73(6), 219-224.
- Shields, C. (2023). Opinion: Don't ban ChatGPT, use it as a teaching tool. *Education Week*.
<https://www.edweek.org/technology/opinion-dont-ban-chatgpt-use-it-as-a-teaching-tool/2023/01> (accessed 2024.09.07)

専門留学における期待・不安要素について
—保育専攻の大学生に対する事前事後の調査から—¹⁾

**Expectations and Anxieties Regarding A Specialized Study Abroad Program:
A Pre-and Post-Survey of University Students
Majoring in Early Childhood Education**

加藤 あや美（桜花学園大学）

内田 政一（桜花学園大学）

ランズベリー ローレン（愛知大学）

キーワード：専門留学、長期必修留学、期待要素、不安要素、事前授業

Abstract

In 2023, 42 Japanese students returned home following a long-term “study abroad” experience in Australia. Since the establishment of the Global Early Childhood Education department, this was the first group to successfully complete their “study abroad” program without being affected by the COVID-19 pandemic and enforced government closures. While in Australia, their goals were to obtain a childcare license, and to improve their English skills. Using the data collected in pre- and post-“study abroad” questionnaires, the students’ expectations, anxieties, and experiences were investigated. This paper presents the results and discusses how the students’ experience differed from their expectations, what anxieties they experienced, what they learnt, and how they felt the experience changed them.

1. はじめに

新型コロナウイルス感染症は 2019 年末に感染者の報告がされてから、わずか数ヶ月ほどの間にパンデミックと言われる世界的な流行となり、世界中に多大なる影響をもた

らした（国立感染症研究所, 2020）。感染拡大を防ぐために、多くの国がロックダウンを行い、海外への渡航には制限がかかった。日本人の海外留学者数は年々増加傾向にあったが、新型コロナウイルス感染症のパンデミックをきっかけとし、留学者数は 98% の大幅な減少となった。一気に落ち込んだ日本人の留学者数であるが、コロナ禍を経て 2021 年度から徐々に増加に転じ、2021 年度は、10,999 人（対前年度比 9,512 人（639.7%）増）、そして 2022 年度で 58,162 人（対前年度比 47,163 人（428.8%）増）となった（文部科学省, 2024）。図 1 は短期留学・語学留学など単位取得を伴わない留学も含めた大学生等（大学生、大学院生、専門学校生、高等専門学校生）の 2009 年度からの留学生数を示したものであるが、留学期間が 3 ヶ月以上の中長期の留学者数については、コロナ禍前の 2019 年度の数字の約 90% となっており、比較的早期に入国制限の緩和に動いた欧米諸国への留学が 2021 年度以降回復に向かい、2022 年度はアジア諸国についても入国制限が緩和され、留学生数の増加につながったと考えられている（文部科学省, 2024）。

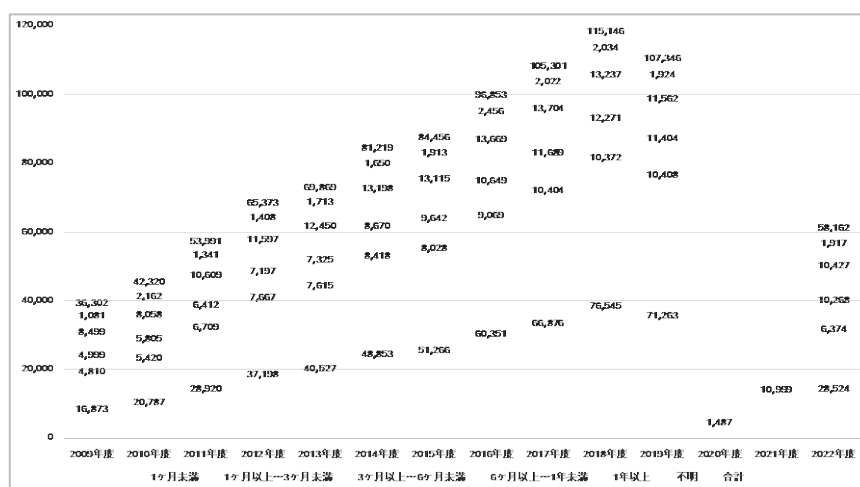


図 1. 大学生の留学者数の推移

上記の通り、2021 年度から留学者数の回復が見られてきており、今後はコロナ禍前のように活発に留学に参加する大学生が増加してくると考えられる。また、留学プログラムの形態は、昨今多様化してきており様々な切り口によるプログラムが運営されている。

本稿は、コロナ禍からコロナ禍後にかけて実施したオーストラリアへの約 11 ヶ月にわたる長期専門留学に焦点を当てて、その留学に参加した大学生が留学に対してどのようなことに期待し、どのようなことを不安に思っていたのか、またそれらの期待と不安

は留学前と留学後で同様のものであったかということについて明らかにするものである。また、語学留学ではなく専門的な分野を海外で学ぶ専門留学ということについても触れ、語学留学とは異なる留学参加者の期待や不安についても考察する。留学前・後の期待や不安の要素を検討することで、留学をより有意義なものとするための事前授業実践の糸口が掴めるのではないかと考えている。

2. 日本における留学プログラムの多様化

2.1. 海外留学の形態

日本において大学生が参加する留学プログラムといえば、語学習得を主な目的とする語学留学が一般的であろう²⁾。現地の語学学校に通い、自分に合ったレベルのクラスで集中的に語学の学習をするものである。多くの大学においても、夏季や春季の長期休暇期間に様々な語学留学プログラムが提供されており、そのプログラムに参加することで一定の単位取得が可能となる場合もある。このような語学留学は3ヶ月未満程度の比較的短期間のものが多く、図1の大学生の留学者数の推移からも見てわかる通り、留学者数10万人を超えたコロナ禍前の2017年度から2019年度は3ヶ月未満の留学者数が全体の70%以上を占めていることから、大学生が最も参加しやすい留学期間であることがわかる。しかしながら、昨今では、短期間の留学プログラムに加えて、中期間を費やすセメスタープログラムや在籍大学と交換協定がある大学に留学する交換留学、単位互換制度などが利用でき単位認定が可能な認定留学など、多様な留学のタイプが生まれている。このように留学の形態が多様化している背景には、留学に対して付加価値を求める傾向があるからではないかと考える。短期間の語学留学は、語学力の条件もあまり厳しくなく、費用面も抑えることができるため、参加者が多い傾向にあるが、現地での滞在が短いため学習期間が限られており大幅な語学力の向上という側面から考えるとそれほど期待を高く持つことはできない。また、翻訳機能やAI技術の開発、それらの実用化等のテクノロジーの進化により語学を習得するだけでは、将来的に活かせる十分なスキルとはいえなくなっている。そのため、留学において、語学力の向上以外の知識や

技能を身につけることができるプログラムへの需要が高まってきている。

文部科学省が官民協働のもとで展開する留学促進キャンペーンである「トビタテ！留学 JAPAN」においても語学だけでなく多様な学びを後押ししている。「新・日本代表プログラム」の目的は、以下の3点を挙げている（文部科学省, 2024）。

- ・日本の未来を作る“グローバルリーダー像”と留学を通じた学びを刷新し、日本の社会課題解決や産業創造に貢献する人材の育成を行う
- ・5年間で5,000名程度以上の学生・生徒に経済面を中心に留学支援を行い、留学機会を創出する事を通じて、日本の留学機運を再度盛り上げる
- ・新たな留学のあり方（ロールモデル）を社会に提示し、日本全体の留学のアップデートを図る

上記3点目にもあるように、「新たな留学のあり方」を示すことを目的の一つとして挙げていることがわかる。また、「新・日本代表プログラム」の特長の一つとしてインターンやボランティアも対象となることが提示されており、座学だけでなく多様な学び方ができるとしている。このように、留学の形態は年々多様化してきており、その背景には留学に付加価値を求める傾向が高まってきているからと考えられ、将来活かすことができる知識や技能を身につけることができる留学プログラムが求められている。

2.2. 専門留学

前節では多様化する留学の形態について述べたが、本節では近年注目を浴びている専門留学について見ていく。専門留学とはどのようなものかという確立された定義は見当たらないものの、上記で紹介した「トビタテ！留学 JAPAN」においては、「専門留学」というカテゴリーが存在し、スポーツ、芸術、調理、技術等を学ぶものという補足的な説明が加えられている。また、留学の多様化が進展する中であって留学を専門とする取り扱い業者のウェブサイト等では「専門留学」という用語を数多く確認することができる。それらの中では、「専門留学」は一般的に海外で特定の分野における専門的な知識や技能を学ぶもので、資格取得やキャリアアップを目指しつつ、より実践的、専門的に学ぶス

タイルであることが述べられていることが多い。そのため、「専門留学」とは自らが専攻する分野における専門的な知識や技能を身につけるために、海外において実践的な学びを深化させるものと捉えることができる。また、語学力の向上に加え、専門分野における資格取得等の付加的な価値が得られる留学であるともいえるものである。

2.3. 大学生が持つ留学に対する期待と不安

黒宮他（2016）では、留学に対する大学生の「不安感」や「期待感」が留学の事前指導を通じ、どのように変化するかということについて検討している。その結果、「不安感」については「漠然とした不安」から「現実的、手段的な不安」へと変化が見られたとしており、特に不安との意見が多かった現地での修学のギャップについてあげれば、留学先の大学の情報を学生が早期から自主的に探索し、事前準備を促す必要性を示していると述べている。また、「期待感」の内容については、語学力の向上といった点のみならず、「自らの人間的な成長」、「視野を広げること」、「自立する」、「将来への選択肢を増やす」といった「自己成長」の位置づけとして「留学」に期待している点が大きいとの指摘をしている。

続いて、岩城（2014）においては、渡航前、渡航中、渡航後において「期待していること」、「不安なこと」等について調査がなされている。渡航前は、全員が「語学・コミュニケーション」の能力向上を期待していたとのことである。「文化」、「健康」、「犯罪」、「環境」においては不安を持つ大学生が多かったと述べており、期待も大きい反面、漠然としか想像できない留学先での生活にほとんどの学生が不安を抱いていたとしている。渡航中については、語学の問題に直面する学生、ルームメイトに関する問題についての言及が多かったとしているが、語学についてはできないことがバネになり、もっと頑張ろうとその先を見据えている学生が大半だったと述べている。さらに、渡航後は全員の学生が行って良かったと感じており、多かれ少なかれ、留学中に何らかのトラブルは抱えるものの、留学自体を振り返る際には、それらすべてが成長の過程であったと捉えていることがわかったとしている。

どちらの調査においても「語学」の能力向上に対して期待を抱く一方で、同時に不安も持ち合わせているということがわかる。また、現地で生活する未来の自分について漠然としかイメージができないことから、「生活全般」について不安があることも窺える。留学への期待は、「語学」だけでなく問題を解決し、乗り越えた後の自分の成長という部分であるということが推察される。また、昨今の留学においては円安の影響を受けてコロナ禍前よりも多額の費用が必要となるケースが増加している。そのため、コロナ禍後に留学を目指す学生にとっては「費用」という要素も大きな不安材料になるのではないかと考えている。

3. 研究の概要

3.1. オーストラリアにおける保育留学

国際教養こども学科は、2018年に設置された保育系学部学科では初の必修長期留学をカリキュラムに組み込んだ学科である。グローバル時代の保育・幼児教育に貢献できる人材の育成をコンセプトとし、2年次2月中旬から3年次1月までの約11ヶ月間オーストラリアに留学し、現地の保育士資格である *Certificate III in Early Childhood Education and Care* を取得することを必修としている。約11ヶ月のオーストラリアへの留学は、最初の約3ヶ月間は語学学校に通い英語を集中的に学習し、残りの期間を保育士資格取得コースで学ぶこととなっている。通常、オーストラリアの専門コースを要する専門学校に入学するためには IELTS5.5（語学学校レベルでは Upper-Intermediate レベル相当）が必要となるため、最初の約3ヶ月間の語学学校において、しっかりと英語学習に取り組む必要がある。

3.2. 留学前準備授業

上述のプログラムに参加する学生には、渡航前のセメスターにおいて、留学前準備授業である「Study Abroad Preparation」という科目が必修科目として設定されている。本授業の計画は以下の通りである。

表 1. 留学前準備授業の内容

回数	授業内容
1	オリエンテーション
2	渡航先（学校）発表、学校ごとによるグループワーク
3	旅行取扱業者によるガイダンス①
4	入学関連書類の説明・作成
5	語学コースの説明・提出書類の確認①
6	保育コースの説明・提出書類の確認②
7	留学費用の説明・質疑応答
8	JASSO 奨学金の説明・質疑応答
9	留学準備ワークショップ① 異文化理解
10	参加者からのアドバイス①
11	参加者からのアドバイス②
12	留学準備ワークショップ② キャリアデザイン
13	旅行取扱業者によるガイダンス②
14	危機管理説明会
15	渡航前最終留学説明会・質疑応答

留学前の準備授業として、表 1 のような流れで授業を行った。必修留学という性質上、全員が同じペースで手続きや準備を進めていく必要があることから、学びの一つとして留学に必要な書類の準備・作成等にある程度時間を割く必要があった。その他の部分においては、学生の留学に対する不安を取り除くという観点から、留学経験がある学生や教職員への質疑応答の時間をなるべく多く設ける工夫をした。さらに、現地においても一人で不安を抱えて過ごすことがないよう、渡航先（学校）が同じ学生同士の交流を深められるようにグループでの作業・活動を多く取り入れた。このように、留学前の準備授業においては、学生同士、教職員とのコミュニケーションを通して、不安要素を取り除くような授業デザインとなるよう心がけた。

3.3. 調査協力者及び調査期間

本調査は、前節で述べた留学前準備授業を受講し、約 11 ヶ月間のオーストラリアへの保育士資格取得を目的とした専門留学に参加した大学生 41 名を対象とし、事前調査として留学前である 2021 年 9 月に Google Form を用いて選択肢より回答する形式と自由記述形式の 2 つのパターンによる無記名のオンラインアンケート調査を実施した³⁾。ま

た、留学後の 2023 年 5 月に事後調査として事前調査と同様のアンケート調査を行った。協力者には、研究目的、方法、参加は自由意思であること、また成績評価には一切影響しないことおよび匿名であり個人が特定されないことをアンケート内冒頭部分に示し、さらに口頭にて説明を行い、学生の同意を得た上で実施した⁴⁾。

調査協力者の英検取得級は、2 級 10 名、準 2 級 15 名、3 級 6 名であった。調査協力者 41 名中 31 名が英検 3 級以上を取得していたという結果であった。また、留学前の TOEIC スコアは以下の通りであった。

表 2. 留学前の TOEIC スコア

2021 年 9 月（準備授業開始時） (n=41)		2022 年 1 月（留学直前） (n=41)	
200 点台	8 名	200 点台	3 名
300 点台	18 名	300 点台	10 名
400 点台	11 名	400 点台	13 名
500 点台	2 名	500 点台	11 名
600 点台	1 名	600 点台	2 名
700 点台	0 名	700 点台	1 名
800 点台	1 名	800 点台	1 名
平均スコア	377	平均スコア	467

準備授業を開始した 2021 年 9 月時点では、平均スコアが 377 点であり、300 点台～400 点台が最も多い層であった。また、200 点台も 8 名いたことから英語の基礎力が安定していない学生が少なからず存在していた。留学直前である 2022 年 1 月時点では、平均スコアが 467 点となっており、約 4 ヶ月前と比較すると平均スコアが 90 点上昇するという結果となった。点数の分布をみると、200 点台～300 点台が減少し、400 点台～500 点台が最も多くを占める層となったことがわかる。先行研究でも見られたように、留学前には自らの語学力に不安を抱える学生が多くいる。そのため、保育・幼児教育系の学科ではあるものの、このプログラムの参加者は留学前の 1 年次～2 年次において、週 2 コマの英語の授業を必修科目として位置づけており、英語に触れる機会をできるだけ作るようにカリキュラムが組まれている。また、自宅外においても自律的に英語学習に取り組めるように、動画を使ったオンライン英語学習およびマンツーマン英会話レッスン

が受けられる English Central を導入した。このような取り組みから、個人差はあるものの留学前にある程度の語学力を伸ばすことができたと思われる。

3.4. アンケートおよび自由記述調査項目

本調査では、①留学に対する期待と不安がどのようなものであったか、②留学経験は期待通りものであったか、③留学に対する不安は実際と同じであったかの3点の研究課題について検討すべく、黒宮他（2016）において海外留学に臨む大学生の実態と課題について調査する際に使用された調査項目を参考とし、研究協力者の学生が参加する留学プログラムに沿う質問項目に修正して調査項目を作成した。

表 3. 留学に対する「期待」に関する項目

	全 く 期 待 し て い な い	期 待 し て い な い	期 待 し て い る	非 常 に 期 待 し て い る
① 英語力が向上する	1	2	3	4
② コミュニケーション能力が向上する	1	2	3	4
③ 視野が広がる	1	2	3	4
④ 専門的な知識が身につく	1	2	3	4
⑤ 現地での実習により保育者スキルが向上する	1	2	3	4
⑥ 海外の資格が取得できる	1	2	3	4
⑦ 将来の選択肢が広がる	1	2	3	4
⑧ 現地の文化を体感できる	1	2	3	4
⑨ 新しい友人、人間関係ができる	1	2	3	4
⑩ 適応・対処能力が高まる	1	2	3	4
⑪ 忍耐力がつく	1	2	3	4
⑫ 自分自身を見直すことができる	1	2	3	4

「期待」に関する項目は表 3 の 12 項目からなる。本調査独自の項目は、「⑤現地での実習により保育者スキルが向上する」、「⑥海外の資格が取得できる」の 2 項目となっており、これらはいずれも本調査の調査協力者が参加する留学プログラムの特徴および内容に沿ったものである。「期待」に関する自由記述の項目は、留学前は「留学に期待すること」、留学後は「期待していたことが期待通りであったか」である。

続いて、「不安」に関する項目について見ていく。「不安」に関する項目は表 4 の 8 項目からなる。また、「不安」に関する自由記述の項目は、留学前は「留学において不安に思うこと」、留学後は「留学を振り返ってみて、もっと準備をしておけばよかったことは

何か」、「留学を経て、自分自身が最も成長（変化）したと思う部分は何か」等である。

表 4. 留学に対する「期待」に関する項目

	で ない	全 く 不 安	で ない	あ ま り 不 安	不 安 で あ る	で あ る	非 常 に 不 安
① 英語力に関して	4		3		2		1
② 費用に関して	4		3		2		1
③ 現地での生活に関して	4		3		2		1
④ 留学先の教育レベルと自己の習熟レベルの ギャップに関して	4		3		2		1
⑤ 留学先での人間関係に関して	4		3		2		1
⑥ 帰国後の学習、単位取得に関して	4		3		2		1
⑦ 手続きの煩雑さに関して	4		3		2		1
⑧ 留学先の生活情報に関して	4		3		2		1

上記の調査項目を用いて、大学生が持つ留学に対する期待および不安に関して 4 件法（留学に対する期待は「非常に期待している＝4、「期待している」＝3、「期待していない」＝2、「全く期待していない」＝1、留学に対する不安は「全く不安でない」＝4、「あまり不安でない」＝3、「不安である」＝2、「非常に不安である」＝1）で回答を得た。回答結果を数値化し全体の傾向を提示した上で、自由記述において得られた回答を含めて考察を試みる。

4. 調査結果

4.1. 留学前調査の結果

(1) 「期待」に関するアンケート調査結果

まず、留学前の「期待」に関するアンケート調査の結果について述べる。表 5 は「期待」に関する 12 項目についてまとめたものである。表 5 の数値は、留学に対する「期待」に関する項目において「4（非常に期待している）」と「3（期待している）」を足した数をパーセンテージで示したものである。すべての項目において、留学前は高い期待を持っていることがわかる。特に、「⑤現地での実習により保育者スキルが向上する」、「⑥現地の保育士資格が取得できる」、「⑦将来の選択肢が広がる」は全員の学生が「期待している」と回答しており、特に高い期待を持っていると考えられる。

表 5 留学に対する「期待」(留学前)

調査項目		期待の割合
①	英語力が向上する	95.1%
②	コミュニケーション能力が向上する	97.6%
③	視野が広がる	95.2%
④	専門的な知識が身につく	95.1%
⑤	現地での実習により保育者スキルが向上する	100%
⑥	現地の保育士資格が取得できる	100%
⑦	将来の選択肢が広がる	100%
⑧	現地の文化を体感できる	97.6%
⑨	新しい友人・人間関係ができる	97.6%
⑩	適応力・対応力が高まる	97.6%
⑪	忍耐力がつく	92.7%
⑫	自分自身を見直すことができる	97.6%

自由記述項目である「留学に期待すること」のコメントは、約半数の 51%が「英語力の向上」について述べられており、選択式のアンケート結果と同様に留学で自らの英語力向上に多くの期待を寄せていることがわかった。その他の記述は、「将来の選択肢の広がり」に関するものが 14.6%、「海外での保育現場経験」に関するものが 9.8%、その他の記述が 24.6%であった。

(2)「不安」に関するアンケート調査結果

次に、留学前の「不安」に関するアンケート調査の結果について述べる。表 6 は「不安」に関する 8 項目についてまとめたものである。

表 6 留学に対する「不安」(留学前)

調査項目		不安の割合
①	英語力に関して	95.0%
②	費用に関して	80.5%
③	現地での生活に関して	80.5%
④	留学先の教育レベルと自己の習熟レベルのギャップに関して	80.5%
⑤	留学先の生活情報に関して	80.5%
⑥	渡航の手続きの煩雑さに関して	70.8%
⑦	留学先での人間関係に関して	68.3%
⑧	帰国後の学習・単位取得に関して	63.4%

表 6 の数値は、留学に対する「不安」に関する項目において「1 (非常に不安である)」

と「2（不安である）」を足した数をパーセンテージで示したものである。「①英語力に関して」については、「期待」も高い反面、「不安」も大きいということが明らかとなった。

「②費用に関して」、「③現地での生活に関して」、「④留学先の教育レベルと自己の習熟レベルのギャップに関して」、「⑤留学先の生活情報に関して」の4項目については、留学生活に対して大まかなイメージしかできないことから、先行研究でもみられた「漠然とした不安」を抱いていると考えられる。

自由記述項目である「留学に対して不安に思うこと」のコメントは、35.7%が「英語力」についてであり、最も記述数が多かった。こちらも選択式のアンケート結果と同様の結果となり、留学で自らの英語力向上に多くの期待を寄せている反面、自分の英語力が留学先で通用するかどうかという不安を抱えていることが見て取れる。その他の記述は、「費用」に関するものが14.3%、「生活全般」に関するものが14.3%、「人間関係」に関するものが10.7%、その他の記述が25%であった。コロナ禍を経て、特に急激な円安や物価高により留学費用がコロナ禍前と比較して高騰していることから、「費用」に関する不安が挙げられたと思われる。また、コロナ禍後半からの渡航となった留学プログラムであったため、新型コロナウイルス感染を含む健康面についての不安も述べられていた。

4.2. 留学後調査の結果

(1) 「期待」に関するアンケート調査結果

本項では、留学後の「期待」に関するアンケート調査の結果について述べる。表7は「期待」に関する12項目についてまとめたものである。表7の数値は、留学に対する「期待」に関する項目において「4（非常に期待している）」と「3（期待している）」を足した数をパーセンテージで示したものであり、比較しやすいように留学前の数値も示した（下線部が留学後の数値）。留学前の「期待」度から実際の留学経験を経て、「期待」していたことに対して客観的に捉え、現実が数値になって現れたと考える。特に、「①英語力が向上する」の項目に関しては、留学前は95.1%が「期待している」と回答していたが、留学後には63.4%となり期待していたほどではなかったと感じている層が一定数

いると推測できる結果であった。

表 7 留学に対する「期待」（留学前後における変化）

調査項目		期待の割合
①	英語力が向上する	95.1% → <u>63.4%</u>
②	コミュニケーション能力が向上する	97.6% → <u>87.8%</u>
③	視野が広がる	95.2% → <u>92.7%</u>
④	専門的な知識が身につく	95.1% → <u>90.3%</u>
⑤	現地での実習により保育者スキルが向上する	100% → <u>78.0%</u>
⑥	現地の保育士資格が取得できる	100% → <u>92.6%</u>
⑦	将来の選択肢が広がる	100% → <u>90.3%</u>
⑧	現地の文化を体感できる	97.6% → <u>97.6%</u>
⑨	新しい友人・人間関係ができる	97.6% → <u>87.8%</u>
⑩	適応力・対応力が高まる	97.6% → <u>90.3%</u>
⑪	忍耐力がつく	92.7% → <u>85.4%</u>
⑫	自分自身を見直すことができる	97.6% → <u>90.2%</u>

以下は、留学前と留学後の TOEIC スコアを比較したものである。なお、留学後の TOEIC は希望者のみの受験であったため、アンケート調査参加者数は 41 名であるが、留学前と留学後の TOEIC スコア比較対象者は 30 名である。

表 8 留学前と留学後の TOEIC スコア比較

2022 年 1 月（留学前）（n=30）		2023 年 2 月（留学後）（n=30）	
最高点	895	最高点	965
最低点	250	最低点	350
平均スコア	467	平均スコア	544

「①英語力が向上する」の項目に関しては、留学前と留学後では、「期待」の度合いに大きな差が現れたが、表 8 の留学前と留学後の TOEIC スコアの比較においては、留学後の方が全体的に高いスコアを取得することができており、平均スコアが 77 点上昇していた。また、留学前準備授業開始時（2021 年 9 月）に受験したスコアと比較すると留学後の TOEIC を受験した学生は平均で 151 点上昇しており、最もスコアが伸びた学生は 380 点の上昇であった。このスコア比較の結果から、英語力全体は向上していると考えることができるが、留学を経験した学生本人たちは試験の結果やスコアだけでなく、英語を使用したコミュニケーションも含んでアンケート項目の「①英語力が向上する」に対す

る「期待」の度合いについて回答したと思われる。したがって、留学を経験した学生本人たちの感覚としては、留学前に「期待」していた英語力の向上は、留学後において思ったよりも得ることができず、その向上しなかった英語力とは英語を使用したコミュニケーション能力と推察される。

また、「⑤現地での実習により保育者スキルが向上する」というこの留学プログラムの主たる目的の一つとして挙げられるものであり、留学前は全員（100%）が「期待している」としたものであるが、留学後は 78.0%という結果となった。おそらく、積極的に現場実習に取り組めた学生にとっては期待通りのものであったが、指示待ちを含む受動的な姿勢で現場実習に参加した学生にとっては、思っていたほど保育者スキルの向上が感じられなかったのではないだろうか。自由記述で得た回答においても、「色々な国の子どもたちと交流することができた」、「違う国の保育について学ぶことができた」というポジティブな記述がある一方で、「現地での実習中は課題に集中することで頭がいっぱいだった」、「実習園のスタッフや子どもの保護者とのコミュニケーションが難しかった」等、目の前の課題をクリアすることで精一杯であったことが想像できる記述が見られた。このことから、専門的な技能である保育者スキルが向上したという実感を得られるには至らなかったことも留学前の「期待」の度合いとは異なる結果となったのではないかと考えられる。

（2）「不安」に関するアンケート調査結果

次に、留学前の「不安」に関するアンケート調査の結果について述べる。表 9 は「不安」に関する 8 項目についてまとめたものである。表 9 の数値は、留学に対する「不安」に関する項目において「1（非常に不安である）」と「2（不安である）」を足した数をパーセンテージで示したものである。比較しやすいよう留学前の数値も示した（下線部が留学後の数値）。

一般的に、留学前の「漠然とした不安」は、留学中に経験を積むことにより軽減されているものであるが、本調査においてもそれとほぼ同様の結果が現れた。留学前は大きな「不安」として挙げられていた「①英語力に関して」であるが、自由記述のコメント

からも「最初は英語力がなくて不安だったが、徐々に慣れてきて不安は薄らいだ」、「相手が理解してくれようとしてくれて、自信を持って会話できるようになった」等、最初は自身の英語力が通用するかどうか不安だったものの、実際のコミュニケーションを通して思ったよりやり取りができることを実感し、「不安」が払拭されていったようであった。

表 9 留学に対する「不安」（留学前後における変化）

調査項目		不安の割合
①	英語力に関して	95.0% → <u>69.1%</u>
②	費用に関して	80.5% → <u>85.7%</u>
③	現地での生活に関して	80.5% → <u>69.1%</u>
④	留学先の教育レベルと自己の習熟レベルのギャップに関して	80.5% → <u>76.2%</u>
⑤	留学先の生活情報に関して	80.5% → <u>57.1%</u>
⑥	渡航の手続きの煩雑さに関して	70.8% → <u>38.0%</u>
⑦	留学先での人間関係に関して	68.3% → <u>61.9%</u>
⑧	帰国後の学習・単位取得に関して	63.4% → <u>69.0%</u>

この留学に対する「不安」の項目において「②費用に関して」と「⑧帰国後の学習・単位取得に関して」は留学経験を積んでも「不安」を軽減することができなかったという結果が示されている。とりわけ費用に関わる自由記述のコメントでは、「費用に関して心配だったが、心配しすぎることはなかった」、「物価の違いなどもあり不安だったが、次第に慣れていった」等の費用に対する「不安」が少しずつなくなっていったという記述も見られるが、「何にどれくらいお金がかかるのかわからず不安だった」、「上がり続ける家賃がずっと心配だった」等、日本と現地の物価の違いに戸惑い、また急な値上がりを経験したことで留学経験が長くなっていても「不安」が消えることがなかったのではないかと考えられる。

5. 考察

5.1. 専門留学の成果

専門留学に参加した学生を対象としたアンケート調査の結果から、専門留学を経て手

応えがあったものは「英語力の向上」ではなく「専門性の向上」であったことが明らかとなった。一般的な海外留学においては、やはり「語学力の向上」が留学の成果として挙げられると思われるが、海外留学をした成果として一番に挙がる項目が「専門性の向上」であるという部分が語学留学をはじめとする一般的な海外留学との違いであるといえる。本調査への協力を得た学生たちは全員現地での保育士資格を取得して帰国することができ、また約 11 ヶ月間の留学期間中の 3 分の 2 以上の期間を現地の保育・幼児教育施設での現場実習期間に充てられていることから、専攻している保育・幼児教育分野の知識や技能が身につく、保育や幼児教育現場での実践的なスキルが向上したと実感しているのではないかと考えられる。図 2 は、留学前の TOEIC スコアと専門コースへの満足度を表したものである。

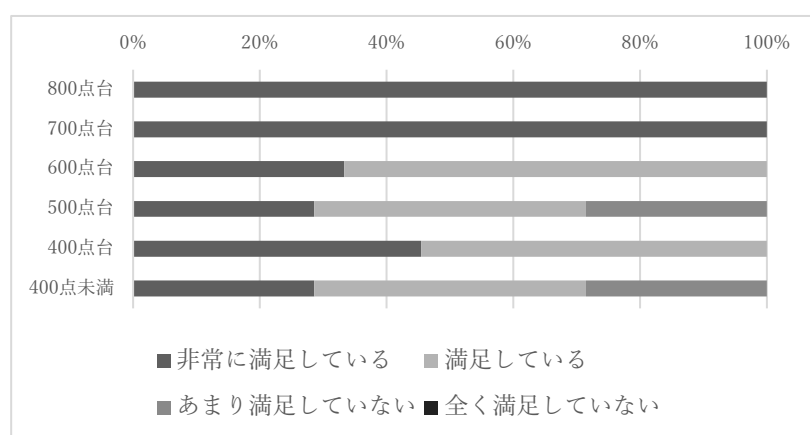


図 2. 専門コースへの満足度と TOEIC スコアの関係

図 2 を見ると、ほとんどが「非常に満足している」もしくは「満足している」と回答しており、この結果から留学前の TOEIC スコアの高低はあまり関係がなく、留学前は英語力が高なくても専門留学に参加した場合は高い満足度が得られるということが示唆される。

一方で、留学前に大きな「期待」を抱いていた「英語力の向上」は思ったよりも伸びなかったという印象を持つという結果となった。これは、前章で述べたとおり、TOEIC スコアは上がっており、ある程度の英語力は向上しているものの、英語を使用したコミュニケーション能力の向上において十分ではないと認識していることから、このような

結果が現れたと推察される。

留学後の自由記述項目「留学を経て、自分自身が最も成長（変化）したと思う部分は何か」では、「忍耐力と精神力」、「視野の広がり」、「行動力の向上」、「コミュニケーション能力の向上」等が挙げられた。この部分から、専門留学に参加し海外での生活経験により「精神面の強化」が成果として挙げられると思われる。海外での生活経験がほとんど皆無な状況から、1年近くの期間を海外で過ごすということは言うほど容易なことではなく、日頃は感じないストレスの中、生活をしていくこととなる。また、専門留学という性質上、現場での実習が伴うことから学校関係者以外の人々との交流が不可欠であり、ローカルなやり取りが飛び交う場所に身を置く必要がある。これらのことが、学生たちの「精神面の強化」という実感につながる要素となったのではないかと考える。

5.2. 専門留学に対する事前準備

本調査の対象とした専門留学のための留学前準備授業においては、約11ヶ月におよぶ長期間の留学に臨む学生たちの「不安」をできるだけ軽減すること、それに付随して困った際にはお互いに助け合える関係づくりをすること等に焦点を当てた授業デザインを考案し、実践した。また、先行研究でも「不安」要素の主たるものとして指摘されていた英語力の向上をサポートするコンテンツも同時に導入した。このような工夫により、留学直後には大きな「不安」として挙げられていたことが、現地に慣れていったことも要因として考えられるものの、留学が進むにつれて薄らいでいったことがわかった。図3は、自由記述にて回答を得た「留学を振り返ってみて、もっと準備をしておけばよかったことは何か」という質問項目とTOEICスコアの関係を示したものである。

図3を見ると、TOEIC800点台を除いて、すべてのレベルにおいて「もっと準備しておけばよかったこと」として「英語学習」を挙げている。参加学生のうち、65.8%が留学前には「英語学習」に取り組んでおくべきというコメントを挙げており、留学前において「英語学習」に取り組んでいたつもりであったが、十分ではなかったとの声が多かった。

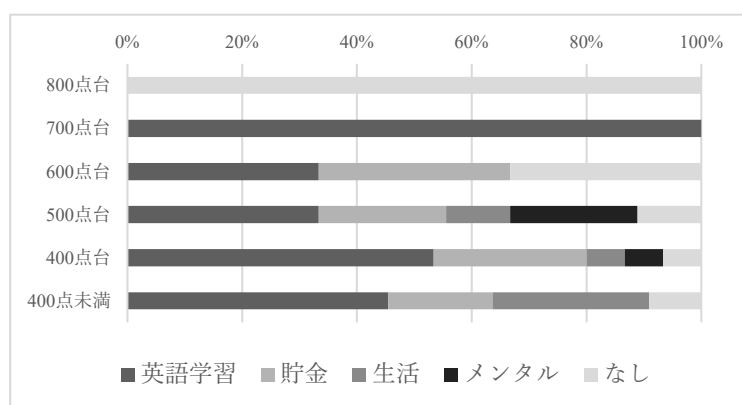


図 3. 専門留学の準備と TOEIC スコアの関係

また、「英語学習」の中でも「スピーキング」のトレーニングが足りていなかったというコメントも挙げられていたが、それよりも「語彙」習得の重要性を述べたコメントが多く見られた。これは、専門コースに留学したことから、多くの保育・幼児教育分野に関連する専門用語が頻出した可能性が高く、それらの知識が不足していたことから専門分野における「語彙」習得の必要性を感じたのではないかと考えられる。これらの結果を踏まえて、留学前準備授業においては、「英語学習」における心構えやその準備、学習習慣の確立をサポートするようなコンテンツの工夫が必要になると考える。また、留学前準備授業だけでなく、カリキュラムに設定されている語学授業（英語）と連携を図り、留学前には専門用語や専門的な文章を読解するトレーニングを実践するといった取り組みも行うことが有効なのではないかと考える。

また、図 3 からは、「英語学習」に続いて「貯金」が挙がっており、経済的に余裕を持って留学をするべきとの指摘がなされていた。すでに述べた通り、近年の急激な円安や物価上昇の影響が非常に大きく、留学費用に関してはコロナ禍前と後では大きな差がある。そのため、留学先の物価等の状況を事前によく調べ、それに対応できる準備をしておく必要があると考える。留学前準備授業や保護者説明会を通して、日本とは異なる経済的な側面について具体例を含めながら丁寧に伝えていく必要があると思われる。

6. おわりに

本稿では、オーストラリアへの約 11 ヶ月間の専門留学に参加した学生を対象とし、留学に対する「期待」と「不安」がどのようなものであるか、またそれらは留学前と留学後でどのように変化するかということについて、アンケート調査をもとに検討してきた。

「期待」の要因は、「専門性の向上」という専門留学ならではの結果となった。「不安」の要因は、「英語力」と「費用」が挙げられ、今後の留学準備授業における課題として浮かび上がった。「英語力」においては、専門留学という観点から専門用語の準備やローカルなやり取りに対応するためのアウトプットの強化等に重点をおく必要がある。また、「費用」については、学生のみならず保護者とも情報を共有し留学先の様子を把握できるように、これまで以上に連携をとることが重要となってくる。

今後ますます多様化していく留学に対して、学生が「期待」する内容を可能な限り達成し、「不安」を軽減するために、留学先の状況を見据えつつ、留学前準備授業のコンテンツを柔軟に検討していく必要がある。

注

1. 本稿は 2023 年 11 月 4 日（土）に開催された JALT Study Abroad SIG Conference において「海外留学における期待・不安要素について～大学生に対する事前事後の調査から～」として口頭発表した内容にデータを追加し、加筆修正を加えたものである。
2. 文部科学省「学生の海外留学に関する調査 2024」の参考資料集によると、高校生の留学したい理由として語学力の向上を挙げる人が最も多いことが示されている。
3. 本研究において行った調査で対象とした留学プログラムは 43 名が参加したが、調査への協力を得た学生は 41 名であった。
4. 本研究は「桜花学園大学人を対象とする研究倫理審査委員会」において承認され、実施したものである。

参考文献

- 岩城奈巳（2014）「渡航前、渡航中、渡航後の振り返りから考える交換留学に対する意識調査」『名古屋大学国際教育交流センター』創刊号, 27-32.
- 国立感染症研究所（2020）「東京都での新型コロナウイルス感染症（COVID-19）の流行（2020 年 1 月～5 月）」<https://www.niid.go.jp/niid/ja/diseases/ka/corona-virus/2019-ncov/2502-idsc/iasr-in/9818-486d01.html>（情報アクセス日: 2024 年 10 月 4 日）
- 黒宮亜希子・橋本由紀子・金沢真弓（2016）「海外留学に臨む大学生の実態と課題について-学生を対象とした調査を基に-」『吉備国際大学研究紀要（人文・社会学系）第 26 号, 121-133.
- 文部科学省（2024）「トビタテ！留学 JAPAN」<https://tobitate-mext.jasso.go.jp>（情報アクセス日: 2024 年 10 月 4 日）
- 文部科学省（2021）「令和 3 年度高等学校等における国際交流等の状況について」https://www.mext.go.jp/a_menu/koutou/ryugaku/koukousei/20230403-mxt_kouhou02-1.pdf（情報アクセス日: 2025 年 2 月 11 日）

査読委員（アルファベット順・2025 年 3 月現在）

馬場	千秋	（帝京科学大学教授）
福島	玲枝	（畿央大学准教授）
林	千代	（元国立音楽大学教授）
飯田	毅	（同志社女子大学特任教授）
木村	友保	（名古屋外国語大学名誉教授）
小西	瑛子	（常磐大学助教）
工藤	泰三	（名古屋学院大学准教授）
森	明智	（愛知工科大学准教授）
佐藤	雄大	（名古屋外国語大学教授）
吉枝	恵	（愛知淑徳大学講師）

編集後記

JACET 授業学ジャーナル 第 5 号が完成いたしました。今回、研究論文 5 本、実践報告 6 本、計 11 本の応募がありました。厳正なる審査の結果、研究論文 2 本、研究ノート 1 本、実践報告 5 本、計 8 本が採択されました。うち 2 本は種別の変更をお願いしました。掲載については、研究論文、実践報告の順とし、それぞれの種別内においては、投稿順としております。

関東、中部、関西の 3 支部の授業学研究会では、本ジャーナルの発行だけでなく、授業学研究大会を毎年オンラインで開催しております。是非、ご発表いただき、その内容をジャーナルにご寄稿ください。

(C.B.)

JACET 授業学ジャーナル 第 5 号

2025 年 3 月 31 日 発行 （出版地：東京都足立区千住桜木）

編集兼発行：一般社団法人大学英語教育学会 授業学研究会（関東・中部・関西）

代表：佐藤 雄大（名古屋外国語大学）

編集委員長：馬場 千秋（帝京科学大学）

