Productive Skills of English A2 to B1-Level CLIL Students

Satomi Fujii
Juntendo University

Reference Data:

Fujii, S. (2024). Productive skills of English A2 to B1-level CLIL students. Growth Mindset in Language Education. JALT. https://doi.org/10.37546/JALTPCP2023-09

This study examines the productive skills of two groups of Japanese EFL undergraduates in a Content and Language Integrated Learning (CLIL) course in Japan and investigates the relationship between these skills and their other L2 performances. Fujii (2022, 2023) reported that one of these two groups of students' proficiency test (i.e., the TEAP test) scores was at A2, and the other at B1; however, their average scores and score ranges of Nation and Beglar's (2007) written vocabulary size test were similar. To understand why the two groups demonstrated different proficiencies, written and spoken productions of four high-achievers (two in the A2 group and the rest in B1) with the knowledge of more than 4,000 word families were investigated. Their final papers and recordings of their presentations and in-class discussions were examined. The findings showed that the B1 students displayed higher discourse competence and experimented more with various linguistic items and ideas in L2.

本研究では、日本のEFL大学生２グループのCLIL（内容言語統合型学習）コースにおける言語産出スキルに着目し、言語産出スキルとその他のL2パフォーマンスとの関係を調査する。Fujii（2022, 2023）は、この２グループのうち一方の学生の習熟度テスト・スコア（TEAPを利用）はA2レベルであり、もう一方はB1レベルであったが、Nation and Beglar (2007)の英語語彙サイズ・テストの平均スコアとスコア幅は２グループ同程度であったと報告している。なぜ２グループが異なる習熟度を示したのかを理解するため、4,000単語家族以上の知識を持つ４名の高得点者（各グループ２名）の学期末のレポート及び発話データを収集し分析した。結果、B1レベルの学生の方が目標言語を用いてより高い談話能力を示し、多様な言語項目や考えを探究及び産出することを試みるということがわかった。

The present study looks at two groups of Japanese EFL undergraduates' productive skills in a CLIL course in Japan and investigates the relationship between these skills and their other L2 performances. Fujii (2022, 2023) reported that one of these two groups of students' proficiency test scores was at A2, and the other at B1; however, their average scores and score ranges of Nation and Beglar's (2007) written vocabulary size test (hereafter, VST) were similar. To understand why the two groups demonstrated different proficiencies, written and spoken productions of students are investigated.

Previous Research

Previous studies that looked at the relationship between university students' receptive VST scores and other skills in the 2000s have shown that the former are good predictors of students' academic performance. Ünaldi & Yüce (2020) demonstrated that students' scores correlated significantly with their grammar proficiency and critical thinking skills, the latter especially so on inquisitiveness and confidence in reasoning. Rahman (2020) found that students' scores were a better predictor of cumulative grade point average (CGPA; that is, overall academic achievement) than English language proficiency at an English-medium instruction (EMI) university in Malaysia. (The proficiency was derived from the students' Malaysian University English Test scores.) Milton and Treffers-Daller (2013) described how, in the UK, monolingual English-speaking students' scores were found to correlate significantly with their scores in assignments and examinations. Additionally, those with larger VST scores obtained high degree classifications. Wang and Treffers-Daller (2017) reported that students' scores correlated with their listening comprehension scores.

Other studies looked at students with similar vocabulary size and reported on variations in the students' performances. Qian and Lin (2020) looked at listening scores of students with similar vocabulary size and found that there was a considerable variation. They claimed that this is due to a) "the fleeting nature of spoken language,

which demands the activation of a short-term working memory" (p. 70), and b) the development of a learner's proficiency level. Fujii (2022, 2023) found score variations in learners' language proficiency test scores, especially in productive skills scores. Albrechtsen (2008) focused on L1 and L2 writing skills of Danish learners and found score variations. She reported that this is due to the students' different writing processes. Some students spent a longer time planning, put a lot of effort into organizing their texts, and/or had good problem-solving skills while writing. These steps were handled insufficiently by lower scorers of the same writing tasks. Albrechtsen (2008) concluded "[it] takes more than a high level of L2 language proficiency [e.g., vocabulary size and reading skills] to operate a good writing process in L2" (p. 152). A qualitative analysis (of the verbalizations of three students in completing their writing tasks) showed that the higher scorers' skills were similar to those in working on L1 writing tasks. They also spent a longer time planning and/or were more in control of the task (i.e., staying on topic or coming up with more elegant solutions) than the lower scorer. They identified the word that tallied with an intended meaning. The lower scorer tended to say "never mind" (p.152) when faced with problems, lacked consideration or did not put in so much effort to find solutions to problems. The proficiency is likely to have forced the students "to settle for less" (p. 151). Participants' personality and/or working methods are also important factors.

Research Questions

This study looks at productive skills of Japanese EFL university students attending a CLIL course in Japan. To understand characteristics of A2 and B1 students' productions, both their written and spoken texts were collected during the academic year.

RQs. How are A2 and B1 students' following texts different: a) planned written texts, b) planned spoken texts, and c) spontaneous spoken texts?

Research Methods

The course and the data collection schedules are shown in Table 1 (see also Fujii, 2021, 2022, 2023).

Table 1

Course and data collection schedules

	Course		Data collection
	A2	B1	
July 2015			• VST (July 15th) • Consent form
October	People & Technology	Language & Communication	• Recording starts (B1 class)
November			• Recording starts (A2 class)
December			• Final paper draft due • Group discussion test (B1 class)
January 2016			• Group discussion test (A2 class) • Final presentation • Final paper due

The participants attended the present English lessons for one academic year, from April to January, at the same university in Japan. They took two proficiency tests: the listening and reading sections of the Test of English for Academic Purposes (TEAP) in April, and all four skills sections of TEAP in January. The first TEAP scores were used to place the students in one of the elementary to advanced EFL courses. In the A2 course, 22 freshmen majoring in either Science or Economics attended, and in the B1, 19 majoring in History, French Literature, Law or Journalism.

In the first semester (April to July), they attended an English for Academic Purposes (EAP) course and took the VST in July. They also received explanations on the present study, and those who agreed to participate signed the consent form. An institutional approval was also obtained.

After the summer break, they attended a soft CLIL course in the second semester (October to January). In Soft CLIL lessons, the lesson aim is geared towards study of language rather than of a specific content or subject, and that several different topics can

be taught in the target language (Ikeda, 2011). A theme selected for the A2 course was *People and Technology*, and the B1, *Language and Communication*. In the former, some topics covered were genetically modified food and robotics, while, in the latter, culture shock and third culture kids. The main textbooks used were those by language learning publishers (e.g., Solorzano & Frazier, 2009).

Classroom interactions in both courses were audio-recorded in October and November. Each participant's draft of the final paper was collected in December. In the A2 class, the students wrote a five-paragraph paper, while in the B1, four to six-page (approximately 1,000 to 1,500 words), both on a topic related to each CLIL course. Their final papers were collected in January. The students' drafts and papers were graded, scanned and returned to each student before the end of the second semester.

The students' 10-minute group discussion and five-minute final presentation audio-recordings were collected in December and January, then later transcribed for analysis. (The participants' names were not removed during the analysis; however, they were removed when the written texts were analyzed by one of the final paper raters. The details are discussed below.) The final presentation topic was the same as the final paper; however, the students were allowed to add new information to and/or delete some information from it. Each student presented twice for two different groups of audience. The second performance is analyzed in this study as they tended to speak more fluently and confidently in the second session.

Two discussion topics were presented on the day of their group discussion test, one of which had been created by the students themselves, and the other, by the instructor. In the A2 class, they were "Q1. What did you do during the winter vacation?" and "Q2. University students do not really need SNS, smartphones or the Internet. Do you agree or disagree? Why do you think so?" In the B1 class, they were "Q1. What do you plan to do in the winter vacation?" and "Q2. You are searching for the best gift for the author or the main character of the book. What would it be? Explain the reasons." In class, each student had compared a book of his or her choice with its movie version and gave a short presentation a few weeks before the discussion test. The aim of the activity was to compare linguistic styles used in books with those in films. The students at both levels chose to talk about their vacation in the first question. Although this was not a CLIL topic, it was accepted because it allowed them to relax and build rapport with their classmates before the main discussion topic. The test was carried out in a CALL room where each student wore a headphone with a microphone placed in front of the mouth. The instructor clicked the record button and recorded all groups' discussion at once.

In each course, two students' performances were analyzed. These four students were selected because their VST scores were above 40 out of 60, which means that they had the knowledge of 4,000 most frequently used word families in English. The A2 students scored 43 and 42, and the B1, 43 and 40. (The rest of the students scored below 40.)

The instructor for the two courses was a Japanese female, a language teacher affiliated to the institution's language education research center. She attended faculty development sessions on CLIL, where CLIL experts gave lectures on the teaching approach and/or sample CLIL lessons. She used English in class, except for a few cases where she provided L1 counterparts of English words to help the students' understanding. She conducted the lessons following the principles and features of CLIL (cf. Ball, Kelly, & Clegg, 2015; Ikeda, 2015). The students tended to use their L1 in communicating with their classmates but switched to English in talking with the instructor and working on speaking tasks in class.

To analyze the written and spoken texts, the Compleat Lexical Tutor (Lextutor) Frequency analysis was used to check a frequency summary of each written or spoken text. In assessing the quality of the student's written production, Wang and Xie's (2022) discourse competence (DC) diagnostic rubrics and the essay evaluation rubrics were utilized. Using these rubrics, two EFL language instructors with more than 15 years of experience teaching at universities in Japan assessed the students' final papers. The DC diagnostic rubrics consist of five components which are further broken down into 10 features: *Topic building* (F1 *Topic/Focus*, F2 *Thesis statement* and F3 *Controlling idea*), *Global coherence* (F4 *Introduction* [or *Reader orientation*], F5 *Body paragraphs*, and F6 *Conclusion*), *Local coherence* (F7 *Theme-rheme* [or *given/new*] *development*), *Logical connectives* (F8 *Connective complexity* and F9 *Connective accuracy*) and *Reader-writer interactions* (F10 *Complexity of hedges and boosters*). Each feature is assessed on a 5-point scale, in which 5 is excellent and 1 is poor.

The essay rubrics include five components, out of which the following two are used in the present study: *Critical Analysis, Argument, Justification and Conclusion* and *Writing* (e.g., free from grammatical errors). The rest of the components were related to the subject covered in Wang & Xie's (2022) research site (e.g., *Demonstrates Understanding of Culture/Business Dilemma*); hence, they were not used in the present study. Each feature is assessed on a 7-point scale, in which the highest score (1.0) is comprehensive, and the lowest (0), poor and/or missing.

Results and Discussion

Differences Between A2 and B1 Students' Written and Spoken texts – Planned texts –

Each participant's final paper and final presentation data are shown in Table 2. For instance, Student BL in the B1 course (with the assessed knowledge of 4,300 word families) chose *Media* as his topic. In his paper, he wrote a total of 630 words (270 word types). In presenting the same topic to his class members, he produced 405 words (202 word types). In both tasks, approximately 77 percent of the words were 1K (i.e., basic) words and about seven to eight percent, the Academic Word List (AWL) words (Coxhead, 2000). The AWL words he produced can be found in Table 3, in which words used in the presentation are highlighted. (He used more academic words in the written output.) In supporting his main ideas, he included information on world events and organizations such as the Arab Spring and ISIS. In his presentation, he did not talk about demonstrations, and instead added a Japanese idol's episode.

Table 2

Students' Final Paper and Presentation Performances

	Ss	Topics	Word tokens (types)		1K 2K AWL		Supporting details	Presentations
			Paper	Present.	Paper	Present.		
B1	BL	Media	630 (270)	405 (202)	76.83 6.19 8.25	77.28 5.43 7.16	The Arab Spring, Treaty, ISIS, Demonstrations	- Adds a Japanese idol's episode - Deletes *demonstrations*
	BS	Education	560 (239)	397 (158)	74.82 6.07 5.71	67.91 5.22 3.23	Militarism, Moral, Laws, Reforms	Adds *totalitarian*, *Yamato-damashii*, *colonialism*, *relic* (e.g., uniforms, sailors)
A2	AL	LED	455 (201)	396 (174)	70.99 4.40 6.81	69.98 4.96 5.46	Fluorescent lamp, Energy crisis, Diode, Voltage	Deletes *solar and wind power generations*
	AS	3D Printers	278 (128)	285 (139)	75.54 10.07 4.68	68.77 9.82 3.51	Buildings in China, Mass production, Guns	- Shows videos on AppleTV for about 2 minutes - Deletes *buildings in China*

Table 3

Academic Words Used in the Students' Final Papers and Final Presentations

Ss	AWL [Number of times used] (Presentation words are highlighted.)
BL	administration[3] collapse[1] collapsed[1] committed[1] communication[1] contributed[2] contribution[1] contributions[1] debate[1] demonstration[3] demonstrations[1] enforced[1] evolved[1] impact[2] incident[2] media[23] monitor[1] networking[1] occurred[2] occurrences[1] role[1] security[1]
	channel[1] culture[1] demonstration[1] finally[1] impact[2] instance[2] media[19] networking[1] role[1]
BS	academic[1] achieve[1] affected[2] aspects[2] communicate[1] communicating[1] communication[1] complex[1] conclusion[1] consisted[1] culture[2] ethic[1] ethics[1] globalization[1] individuality[1] issued[1] military[3] percent[1] philosophy[1] prime[1] rigid[1] style[2] technology[1] traditional[1] unification[1] uniform[1]
	conclusion[1] emphasized[1] impacted[1] military[2] revolution[2] topic[1] uniform[4] violate[1]

Ss	AWL [Number of times used] (Presentation words are highlighted.)
AL	chemical[1] compound[1] conclusion[1] constantly[1] consumption[3] converted[1] depression[1] energetic[1] energy[9] exposed[1] formulate[1] generation[2] methods[1] negative[2] processed[1] shift[1] sustain[1] theory[2]
	advocating[1] chemical[1] compound[1] conclusion[2] constantly[1] converted[1] depression[1] energy[5] exposed[1] negative[3] processed[1] sustain[1] theory[2]
AS	collapsed[1] conclusion[1] confirm[1] contrary[1] corporation[3] editors[1] expert[2] guarantees[1] instance[1] job[1]
	co-operation[1] conclusion[1] corporation[2] dimensional[1] expert[1] image[1] job[1] process[1] structure[1]

The B1 students, compared to the A2 students, produced more words in each paragraph (108 and 72 words per paragraph on average, respectively), wrote more objectively, and received higher ratings. Firstly, Students BL and BS produced on average 124 and 93 words per paragraph, respectively, and Students AL and AS, 90 and 54, respectively. Students BL and AL had the same VST score (the knowledge of 4,300 word families), but the former produced slightly more AWL words than the latter, both in written and spoken texts. The following are the first four sentences of each student's paragraph. These paragraphs were selected as it contained the highest percentage of AWL words in each of the students' paper. The AWL words are underlined.

- Student BL: These days, the incident influenced by media is "Arabian Spring". A boy was enforced by government in Tunisia, and he could not work. As result, he committed suicide. A lot of people contributed this incident to Facebook and satellite broadcasting also televised this news (*sic*; followed by nine more sentences containing eight AWL words such as *demonstration* and *collapse*).
- Student AL: Using LED is good choice. Because of population growth, the Earth will confront energy crisis. We should economize consumption of energy. There are some methods, for example, solar power generation, wind power generation and LED (*sic*; followed by one more sentence containing two AWL words *consumption* and *energy*).

The percentage of AWL words was higher in the A2 student's paragraph than in the B1 student:10 percent (14 out of 138) in Student BL's paragraph, and 18 percent (8 out of 45) in Student AL's. In the case of Students BS and AS with lower VST scores than the two students above (4,000 and 4,200 word families, respectively), the B1 student used more AWL words: 10 percent (7 out of 69) in Student BS's paragraph, and six percent (3 out of 53) in Student AS's.

- Student BS: In conclusion, today's Japanese education is consisted of some facts that is too complex. The westernization gave us unification but we lost individuality. The conservatism will not try to reform or change, just holding. Although it was fine in Meiji era, now we are demanded to be assertive (*sic*; followed by two more sentences containing two AWL words *communicate* and *globalization*).
- Student AS: 3D printing was invented in America. At first, it was not used by many people, but in 2009, it was spread and then a lot of people and corporation have been using 3D printing. It is bad situation. We should not use 3D printing unless an expert guarantees that 3D printing is safe (*sic*).

Secondly, the quality of Student BS's paragraph resembles that of Student BL's than the two A2 students. For instance, the papers of B1 members are more objective and closer to texts of newspaper articles and tertiary-level textbooks, but those of A2 members do not reveal such qualities. Lastly, the evaluators rated that the B1 members' writings are more globally and locally coherent, use connectives more complexly and accurately, and are better at reader-writer interaction (Table 4). The B1 members received higher marks on six of the features of the DC diagnostic rubrics (i.e., *Global coherence*'s two features, *Local coherence*, *Logical connectives*, and *Reader-writer interactions*) and the Essay evaluation rubrics. There was a significant difference in the mean scores of 'F4 *Introduction* (or *Reader orientation*)', $t(2) = -6.00$, $p < .05$, with the B1 members receiving higher scores than the A2 members. That is, the introduction section provided a more appropriate background or context to orient the reader.

Table 4

Ratings of the Students' Final Papers

Features	A2		B1		Mean difference	t_{obsAB}
	Mean	SD	Mean	SD	B1-A2	
F1	4.000	0.354	3.500	0.707	-0.500	0.894
F2	3.500	1.061	3.000	0.707	-0.500	0.555
F3	3.750	0.530	2.750	1.061	-1.000	1.193
F4	3.250	0.177	4.000	0.000	**0.750***	-6.000
F5	3.250	0.177	3.250	0.354	0.000	0.000
F6	2.750	0.177	3.000	0.707	0.250	-0.485
F7	3.250	0.177	3.500	0.707	0.250	-0.485
F8	3.500	0.354	3.750	0.354	0.250	-0.707
F9	3.250	0.177	3.750	0.354	0.500	-1.789
F10	2.000	0.000	2.250	0.354	0.250	-1.000
DC overall	3.250	0.212	3.275	0.530	0.025	-0.062
Critical analysis, etc.	0.525	0.035	0.625	0.035	0.100	-2.828
Writing	0.438	0.088	0.600	0.071	0.163	-2.030

The B1 students searched for academically more challenging sources such as the Arab Spring and militarism, and paraphrased the original texts in ways that helped readers understand their ideas. They were skilled at storytelling. The texts of the A2 students also contained interesting pieces of information, such as Chinese buildings made with 3D printers or mechanisms behind LED lights; however, the amount of explanation was limited, long and/or unclear. They were also less coherent, as observed in the ratings. Additionally, during the five-minute presentations, the A2 students stopped speaking and just showed videos for two minutes and/or deleted some of their supporting details worthy of mentioning. The students tended to keep things to a minimum.

Differences Between A2 and B1 Students' Spoken Texts – Spontaneous Outputs –

The students' performances on the group discussion test can be found in Table 5. Student AL produced the largest number of words (n = 427), and the lists of 2K and AWL words he produced also show that his utterances included the widest range of words. He also expanded his group members' opinions. Students BS and AS follow Student AL, with about 320 words, then Student BL, 295. The students with the knowledge of 4,300 word families (Students BL and AL) produced more AWL words than the other two students. However, the B1 students' production of 1K, 2K and AWL words exceeded that of the A2 students (i.e., the former's totals were 79.65 and 74.29, while the latter's, 72.19 and 72.10). About 28 percent of A2 students' words were L1 terms and fillers such as *mm* and *ahh*. Similarly, Student BS often produced off-list words, many of which being L1 terms. Other interesting points to note are that a) the four students volunteered to be a discussion leader during the test and expanded opinions of their group members; and b) Student BL, the most successful performer in the planned outputs became passive, producing the least number of words in spontaneous output. He also misunderstood one of the discussion topics.

Table 5

Students' Group Discussion Test Performances

Levels	Ss	Contents	Word tokens (types)	1K 2K AWL	2K used	AWL used	No. of group members
B1	BL	Keigo Higashino, Atami, etc.	295 (116)	75.25 2.03 2.37	christmas[1] exciting[1] imagine[2] inform[2]	author[6] finally[1]	2
	BS	Meiji Jingu, etc.	319 (131)	72.10 1.25 0.94	football[1] hello[1] searching[1] tall[1]	author[1] job[1] theme[1]	4

Levels	Ss	Contents	Word tokens (types)	1K 2K AWL	2K used	AWL used	No. of group members
A2	AL	Hatsumode & Ekiden, IT neutral	427 (170)	67.52 2.57 2.10	club[1] convenient[1] copper[2] discussion[1] during[1] entertainment[2] information[1] tasting[1] track[1]	convince[1] goal[1] negative[2] neutral[1] physically[1] task[2] topic[1]	4
	AS	Hospitalization, IT necessary	317 (117)	68.65 2.82 0.63	convenient[1] discussion[1] during[1] hello[1] hospital[2] sorry[2] tool[1]	communicate[1] survey[1]	2 (Quiet partner)

Pedagogical Implications

The findings showed that the B1 members had a higher discourse competence and experimented more with various L2 linguistic items and ideas than the A2 members. The latter students tended to settle for less (especially in giving a presentation), which was a similar quality observed in Albrechtsen's (2018) research. However, since the A2 students, when compared with the rest of the A2 course members, had similar VST scores as the two B1 members, volunteered to be a discussion leader, and actively interacted with the instructor, they could have benefitted more if they had been pulled out from their regular class toward the end of the academic year and joined the B1 students. This is because the proficiency level of Students AL and AS was at B1-level in the second semester. Joining B1 members would allow them to find role models and observe the B1 students' performances. The findings also imply that it is essential for each learner to experience different kinds of writing and speaking activities in class, as the quality of a student's outputs varies depending on the skills, as can be seen in Student BL's case (i.e., active and skilled in written performances but passive and less skilled in spoken ones). The students in the present study had the opportunity to notice their weaknesses and tended to comment that they need to improve these areas.

Limitations and Future Research

There are several limitations in the study. One is that both the students' written and spoken outputs were collected toward the end of the academic year. Because the A2 members' proficiency reached that of the B1 learners by then, the data of each student's performance should have also been collected before the end of the year. Another limitation is that this study was conducted in intact classrooms, and, in the case of group discussion test data, for instance, there were only two students in Student BL and AS's groups, while there were four in the other groups. This is likely to have affected the study outcomes, therefore needs to be controlled for in future studies. Additionally, as the students often used their L1, and also, as other researchers have claimed that "it takes more than a high level of L2 language proficiency" in producing written texts (Albrechtsen, 2008), other areas, such as the students' L1 proficiency and critical thinking skills, should be investigated. The data of the rest of the students with the VST scores below 40 who attended the course with the students should be further analyzed to understand the differences between the B1 and A2 classes.

Conclusions

The study showed that the B1-level students' outputs were more sophisticated and coherent than those of the A2-level students with similar VST scores. The former students were also better at storytelling, adjusting the level of their written texts in giving a presentation so that the audience would enjoy and comprehend them better. They included more technical information (e.g., mentioning complicated world events) that they had learned in the present CLIL course as well as on their own. Higher language proficiency is likely to allow learners to explore and experiment with more linguistic items and ideas in L2. Such actions were not seen as much in the A2 members' performances. In future research, factors that lead to these differences should be investigated.

Bio Data

Satomi Fujii is an assistant professor at Juntendo University. She is currently teaching Linguistics, academic skills, and foreign language courses at universities in Japan.
<s.fujii.ip@juntendo.ac.jp>

Acknowledgement and Funding

I am grateful to the editors and the anonymous reviewers for the invaluable advice on an earlier draft. This work was supported by JSPS KAKENHI Grant Number JP21K13059.

References

Albrechtsen, D. (2008). Writing in two languages. In D. Albrechtsen, K. Haastrup, & B. Henriksen (Eds.), *Vocabulary and writing in a first and second language: Processes and development* (pp. 112-159). Palgrave Macmillan.

Ball, P., Kelly, K., & Clegg, J. (2015). *Putting CLIL into practice*. Oxford University Press.

Cobb, T. (n. d.). Compleat Lexical Tutor (Lextutor) [Website]. Unpublished instrument. Retrieved from http://www.lextutor.ca

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, *34*(2), 213-238. https://doi.org/10.2307/3587951

Fujii, S. (2021). Vocabulary learning processes of Japanese EFL university students in a content and language integrated learning (CLIL) course. *The Journal of Institute for Language and Education Research*, *38*, 167-182.

Fujii, S. (2022, November). *Test performances of B1-level university CLIL course students in Japan*. Paper presented at the JALT 2022 International Conference, Fukuoka, Japan.

Fujii, S. (2022). Assessing the effects of CLIL on Japanese EFL university students' receptive vocabulary knowledge. *The Journal of Institute for Language and Education Research*, *39*, 123-136.

Fujii, S. (2023, November). *Productive skills of English A2 to B1-level CLIL students*. Paper presented at the JALT 2023 International Conference, Ibaraki, Japan.

Fujii, S. (2023). Students' initial level of vocabulary and their CLIL course test performances: The case of CEFR A2-level university students in Japan. *The Journal of Institute for Language and Education Research*, *40*, 131-144.

Ikeda, M. (2011). CLILの基本原理[Basic principles of CLIL]. In Y. Watanabe, M. Ikeda, & S. Izumi (Eds.), CLIL（内容言語統合型学習）上智大学外国語教育の新たなる挑戦　第一巻　原理と方法 [*CLIL (Content and language integrated learning): New challenges in foreign language education at Sophia University, Volume 1: Principles and methodologies*] (pp. 1-13). Sophia University Press.

Ikeda, M. (2015). グローバル・リーダーの素養を伸ばすCLIL型授業[CLIL lessons that develop global leaders' knowledge]. 英語展望 [*ELEC BULLETIN*], *122*, 22-28.

Milton, J., & Treffers-Daller, J. (2013). Vocabulary size revisited: The link between vocabulary size and academic achievement. *Applied Linguistics Review, 4*(1), 151–172. https://doi.org/10.1515/applirev-2013-0007

Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, *31*(7), 9–12.

Qian, D. D., & Lin, L. H. F. (2020). The relationship between vocabulary knowledge and language proficiency. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 66-80). Routledge.

Rahman, A. (2020). Receptive vocabulary size as a predictor of undergraduates' overall cumulative grade point average with English as medium of instruction in universities. *Asia-Pacific Social Science Review*, *20*(3), 56–75.

Solorzano, H., & Frazier, L. (2009). *Contemporary topics 1: Academic listening and note-taking skill*. Pearson Education.

Ünaldi, I., & Yüce, E. (2020). The relationship among vocabulary size, grammar proficiency, and critical thinking skills of adult language learners. *Adult Learning, 32*(2), 70-78. https://doi.org/10.1177/104515952095947

Wang, Y., & Treffers-Daller, J. (2017). Explaining listening comprehension among L2 learners of English: The contribution of general language proficiency, vocabulary knowledge and metacognitive awareness. *System, 65*, 139-150. https://doi.org/10.1016/j.system.2016.12.013

Wang, Y., & Xie, Q. (2022). Diagnosing EFL undergraduates' discourse competence in academic writing. *Assessing Writing*, *53*. https://doi.org/10.1016/j.asw.2022.100641