英語語彙サイズと CLIL 授業語彙テスト結果の関連性
―国内大学生（CEFR A2 レベル）のケース―

藤井　里美

Students' initial level of vocabulary and their CLIL course test performances:
The case of CEFR A2-level university students in Japan

FUJII Satomi

要旨

　本稿では、国内大学における内容言語統合型学習（CLIL）を取り入れた英語授業の受講生の「語彙サイズ」および「授業語彙理解度」の関連性を調べた。各被験者には CLIL 授業開始前に Vocabulary Size Test（Nation & Beglar, 2007; VST）を、その後、CLIL 授業で扱われた語彙を含んだ「語彙テスト」（事前および事後テストの計 3 つ）を受験してもらった。VST と語彙テスト間の傾向を見たところ、4,000 単語家族以上の語彙知識を持つ学生は、全語彙テストで平均点以上のスコアを取った。一方、2,000 から 4,000 単語家族の知識を持つ学生のパフォーマンスには一定の傾向は見られなかった。前者の学生は、後者の学生と比べ明確な学習目標を持ち、ネイティブ・スピーカーと接した経験があった。

キーワード：assessments, classroom performance, CLIL, EFL, vocabulary size

Introduction

　The relationship of a learner's vocabulary size to estimated language skill(s) has been reported in many studies; however, the relationship between one's vocabulary size and actual performance is under-researched. Vilkaitė-Lozdienė and Schmitt (2020) pointed out the limitation as follows:

　So far the research on high-, mid-, and low-frequency vocabulary is based mostly on corpus data or on studies of lexical coverage. Coverage figures are mostly drawn from studies where

participants read texts with certain percentages of words missing. However, more empirical evidence from quasi-experimental classroom studies would be useful to see what learners can actually do in real-world teaching contexts. (p. 92)

In this study, L2 vocabulary size and performance of English as a Foreign Language (EFL) university students attending a content and language integrated learning (CLIL) course in Japan are investigated. In CLIL lessons, the target language "is used for the learning and teaching of both content *and* language" (Coyle, Hood, & Marsh, 2010, p. 1). CLIL students also get a lot of opportunities to interact with classmates and are exposed to meaningful, contextualized L2 input (cf. Ball, Kelly, & Clegg, 2015). A CLIL course was selected for this study because CLIL and other similar educational approaches are becoming more and more popular in Japan, and such learning environments seem effective for learners' vocabulary growth. However, not enough research has been conducted in this regard; hence, the effects of CLIL on students' vocabulary knowledge are still unclear. The present study aims to explore this point.

1. Previous Research

Lexical Coverage Study Findings (Up to 9K Word Families)

There have been some important lexical coverage research findings in the last 20 years, as shown in Table 1. Note that a *word family*, counting units mainly used in this paper, includes a term's headword, its inflected forms (e.g., plural or past tense) and closely related derived forms (with affixes such as *-ly* and *–ness*) (Nation, 2001, p. 8). There is ongoing debate among researchers on which counting units to be used (word families, lemmas or others), but most of the studies so far have used word families (Vilkaitė-Lozdienė & Schmitt, 2020, pp. 84-85). There are three other points to note. Firstly, there are two approaches to determining *lexical coverage*; that is, finding out what percentage of a) words in a piece of discourse participants need to know for adequate comprehension, and b) tokens in a text/corpus was covered by particular word lists (Vilkaitė-Lozdienė & Schmitt, 2020, p. 84). For instance, of the studies listed in Table 1, van Zeeland and Schmitt (2013) used the former approach, and the rest, the latter. Secondly, there are many limitations at the moment; for instance, a) studies investigating the relationship between vocabulary knowledge and productive skills are insufficient (Qian & Lin, 2020, p. 71); and b) most of the studies are based on corpus data or participants' receptive performances when exposed to texts with missing words, and not enough classroom studies are conducted to investigate the relationship between the findings and actual in-class performances (Vilkaitė-Lozdienė & Schmitt, 2020, p. 92). Lastly, "the vocabulary size score is likely to be a generous estimate of vocabulary size" (Coxhead, Nation, & Sim, 2015, p. 127).

Table 1

Lexical Coverage Study Findings in the Last 20 Years (Up to 9K Word Families)

Vocabulary size estimates (in word families)	Lexical coverage (%)	Modules (Contexts)	Participants OR Corpus-based	Studies
750 − 2K	90	Listening (Informal narrative by NSs)	L2 English learners	van Zeeland & Schmitt, 2013, p. 475
2K − 3K	95			
2K	95	Daily verbal communication (General spoken discourse by the British and Irish in the 1990s)	Corpus	Adolphs & Schmitt, 2003, p. 430
3K	96			
3K + proper nouns + marginal words + the AWL	95	Academic spoken discourse (Lectures & seminars at two British universities in the 2000s)	Corpus	Dang & Webb, 2014, p. 66
4K + proper nouns + marginal words	96			
8K + proper nouns + marginal words (+ the AWL)	98			
3K + proper nouns	95	Listening (a. Informal daily conversation)	Corpus	Nation, 2006, p. 77
4K + proper nouns	96	(b. A children's movie *Shrek*)		p. 75
6K & 7K + proper nouns	98	(a. Informal daily conversation)		p. 77
7K + proper nouns	98	(b. A children's movie *Shrek*)		p. 75
3K + proper nouns	98	Reading (a. Graded readers *The Picture of Dorian Gray*)	Corpus	Nation, 2006, p. 72
4K + proper nouns	95	(b. Academic texts; newspapers)		p. 72
		(c. Novels)		p. 71
8K − 9K + proper nouns	98	(b. Academic texts; newspapers)		p. 72
		(c. Novels)		p. 71

Written texts. Researchers have claimed that the 95% threshold is considered as the minimal lexical coverage, and the 98% threshold, the optimal lexical coverage, in reading comprehension (Laufer, 2020, p. 1077). Nation (2006) found that at least 4,000 word families (including proper nouns) are needed to understand 95% of newspapers, academic texts and/or novels (pp. 71-72). To reach the 98% threshold, he concluded that the number rises to 8,000 to 9,000 word families plus proper nouns (pp. 71-72). According to Laufer (2020), learners "should acquire sight vocabulary [i.e., a reader has enough knowledge of the term out of context] of at least 5,000, preferably 8,000 word families" (p. 1077) to cope with authentic texts such as academic reading materials. Nation's estimate is that a vocabulary of 8,000 word families entails knowing 34,660 word types (family members) (Nation, 2006, p. 65). (See Vilkaitė-Lozdienė & Schmitt [2020, p. 89] for study findings based on lemmas.)

Spoken texts. An estimate of needing a vocabulary of 2,000 word families for successful spoken discourse comprehension has long been well-known; however, recent findings have shown that more than 2K is likely to be necessary. Adolphs and Schmitt (2003) used a modern spoken corpus called the Cambridge and Nottingham Corpus of Discourse in English (CANCODE) and reported that "2,000 word families made up less than 95 per cent coverage" (p. 425). In their following analysis, they used both the CANCODE and the spoken component of the British National Corpus (BNC) and found that "around 5,000 individual words were required to achieve about a 96 per cent coverage figure" (Adolphs & Schmitt, 2003, p. 425). This means that 3,000 word families cover 96% of native speakers' daily verbal conversation (p. 430). According to Vilkaitė-Lozdienė and Schmitt (2020), learners who know the 3,000 most

frequent word families can communicate in a range of situations, and studies tend to show that these word families cover 95% of written and/or spoken texts (p. 87). For instance, in Nation's (2006) study on spoken texts, 3,000 word families plus proper nouns covered 95% of informal daily conversation, and 6,000 to 7,000 word families plus proper nouns covered 98% (p. 77). In the case of academic spoken discourse at university level, 4,000 word families (excluding proper nouns and marginal words) covered 96% of the texts in Dang and Webb's (2014) study. To reach the 98% threshold, students attending British universities would need a vocabulary of 8,000 word families (plus proper nouns and marginal words) with the help of the Academic Word List (AWL) (Dang & Webb, 2014, p. 66).

Listening comprehension. Van Zeeland and Schmitt (2013) asked their native and non-native participants to listen to recorded informal narrative passages and found that many of their non-native participants' listening comprehension was a) adequate at 90% coverage, and b) relatively good at 95% coverage (p. 457). According to their study, the former coverage would require between 740 to 2,000 word families, and the latter, 2,000 to 3,000 word families (p. 457). Nation's (2006) estimate is that a vocabulary of 3,000 word families (if the aim is the 95% coverage) entails knowing 16,137 word types, and a vocabulary of 6,000 word families (the 98% coverage), 28,015 word types (p. 65). However, as Schmitt (2010) pointed out, "there is simply not enough evidence to confidently establish a coverage requirement for listening at the moment" (p. 7). Comparing reading and listening processes, Qian and Lin (2020) claimed the following important points:

> While similar factors are at work in the process of reading and listening, such as lexical knowledge, genre knowledge, topic familiarity, text organization, and world/subject knowledge, the fleeting nature of spoken language, which demands the activation of a short-term working memory, requires a different role of vocabulary knowledge in the listening process. When reading, learners can refer back to lexical items in the text to facilitate their comprehension, but this is not possible during the fast-moving process of listening, which can be onerous to many learners. This could explain the considerable variation identified in the listening comprehension scores awarded to learners with a similar vocabulary size (e.g., Stæhr, 2009; van Zeeland & Schmitt, 2013) since their general language proficiency levels can be different. However, a counterargument holds that spoken language is usually not as lexically dense as written language. This feature, coupled with the assistance of nonverbal communicative devices, such as tone, gesture, and facial expressions, can alleviate the difficulty in lexical processing, and such facilitation is nonexistent in a typical reading process. (p. 70)

CLIL and Regular EFL Class Students' Vocabulary Learning Processes

Positive effects of CLIL on students have been reported in many of the previous studies, but there are also study findings which have implied that the approach should be implemented with care. (See more in Fujii, 2022.) In the case of the former, for instance, Xanthou (2011) investigated CLIL 6th graders' vocabulary knowledge and found that they outperformed EFL learners. Merikivi and Pietilä (2014) studied CLIL 6th graders' (aged 11 to 12) both receptive and productive vocabulary size and discovered that these learners reached the same level as that of mainstream EFL 9th graders (aged 14 to 15). On the other hand, Reynaert (2019), for instance, looked at students' productive vocabulary size, using Laufer and Nation's (1999) Vocabulary Levels Test, before and after exposure to CLIL and mentioned the following important findings: a) previous experience with CLIL led to better test scores, and b) "significant changes in general productive vocabulary increase are observable after two years of experiencing CLIL education" (p. 158). Castellano-Risco, Alejo-González, and Piquer-Píriz's (2020) summary of studies (conducted between 2009 to 2016) investigating 4th to 10th graders' receptive vocabulary size showed that "after 1000 hours of exposure to English, [both CLIL and traditional EFL] learners are approximately able to understand the 1K band" (p. 3). Gierlinger and Wagner (2016) looked at CLIL learners' receptive vocabulary growth using a vocabulary size test *X-Lex The Swansea Levels Test* (Meara & Milton, 2003) and found the following: a) both the CLIL and non-CLIL (regular English) class groups' (all 13 to 14 year-olds, the majority of them L1 German students studying English in Austria) vocabulary growths were seen after the intervention (p. 48); b) CLIL group outperformed the control group "in terms of absolute test scores, [but] the relative gain of the control group [exceeded] the CLIL pupils by far" (p. 49); and c) the control group outperformed in terms of overall receptive vocabulary growth.

2.　Research Questions

Previous research findings on the relationship between learners' vocabulary knowledge and CLIL course performances are mostly of younger learners living in Europe, and actual classroom performance of EFL undergraduates in Japan is still unclear. In the present study, the following three points are investigated to compare classroom performance of Japanese CLIL university students with higher versus lower initial level of vocabulary.

1. How do performances of Japanese EFL university students with a higher initial level of vocabulary and those with a lower one differ on vocabulary tests in a CLIL course?

2. How many words are likely to be needed at the onset of the present CLIL course for the students to perform successfully on all the vocabulary tests?

3. What are some characteristics of the high- and low-achievers in this course?

3.　Research Methods

　Twenty-two undergraduates (16 males and six females) attended the course, out of which 20 agreed to participate in the study. They were freshmen attending a private university in Tokyo. Their placement test scores (i.e., TEAP listening and reading sections), administered by the institution a few days after their entry into the university, showed that the participants' English skills are at CEFR A2-level. The average score of the two receptive skill sections combined was 88.25 (SD=0.89, low-high: 87.00-89.00) out of 200. They were placed in an elementary-level course, which they were required to attend twice a week for 90 minutes each. The students majored in either Science or Economics. The instructor was a Japanese female, who solely used the target language and conducted the lessons following the principles and features of CLIL (cf. Ball, Kelly, & Clegg, 2015; Ikeda, 2015).

　In this course, the students attended English for Academic Purposes (EAP) lessons in the first semester (April to July, 2015), then Soft CLIL lessons in the second semester (October 2015 to January 2016). The focus of Soft CLIL lessons is said to be more on language than on a content/subject (Ikeda, 2011, p. 10). The main topic of the semester was *People and Technology*, for which its subtopics (e.g., genetically modified food) were selected from the textbook *Contemporary topics 1: Academic listening and note-taking skill* (Solorzano & Frazier, 2009). In July, the students took Nation and Beglar's (2007) vocabulary size test (monolingual, 14,000 version; hereafter VST) up to the 6,000 level. During the second semester, the students took three vocabulary tests (pre-, post-, and delayed post-session; hereafter PreT, PostT and DPostT) which contained the words used in this CLIL course. In PreT (given in October), 91 words selected by the instructor as the target words were included. On PostT (in December), 120 words were tested, 91 of which were the same as PreT, and the rest, words the students had selected themselves (i.e., words they reported they had learned in class after attending the lesson). Lastly, in DPostT (in January), 22 words from PostT were tested. (E.g., see PreT words in Appendix and the rest in Fujii [2021].) All the tests were given announced, and each student was asked to write the meanings of the words in either L1 or L2. Each correct answer was given one point, and incorrect, zero. (For more details, see Fujii [2021, 2022].)

4.　Results

Performances of Students with a Higher or Lower Initial Level of Vocabulary

　The participants' test results are shown in Table 2. The VST score multiplied by 100 equals

to each student's total receptive vocabulary size. Student A is the highest scorer (a vocabulary of 4,300 word families), and Student T, the lowest (2,500 word families). The average VST score was 33.80 (SD=4.48) out of 60. The top two VST scorers, both in the 4K level band, achieved above-average scores on all the tests, while the bottom two scorers (both in the 2K level band) performed poorly on all the tests. Scores of the rest of the learners in the 2K to 3K level bands did not show straightforward tendencies. For instance, the third-highest scorer of VST (Student C with 38) achieved below average on the three tests, while Students D and E with 37 outperformed Student C. Similarly, four others (Students I, O, P and R) outdid Student C on the three tests.

A Vocabulary Size Needed at the Onset of the Present CLIL Course

The students who scored over 40 on VST scored high on all the tests; therefore, those who know the most frequent 4,000+ word families of English are likely to perform successfully in this CLIL course. Since the participants who scored 37 also did well on all the tests, those with a vocabulary of around 3,700 word families may also do successfully. However, a cut-point is hard to determine in the data. More research should be conducted to understand performances of those with VST scores of around 25 to 45.

Table 2

The Students' VST, PreT, PostT and DPostT scores

Students (n = 20)	VST	PreT	PostT	DPostT
A	43	62	87	15
B	42	61	77	14
C	38	45	53	8
D	37	57	84	17
E	37	52	70	11
F	36	43	59	12
G	36	41	70	13
H	35	42	52	10
I	34	72	88	13
J	34	47	63	12
K	34	41	49	6
L	33	47	56	9
M	33	46	69	9
N	33	40	59	6
O	31	72	78	11
P	30	60	98	13
Q	30	45	55	11
R	29	66	85	14
S	26	34	40	7
T	25	32	48	6
Total	60	91	120	22
M (S)	33.80 (4.48)	50.25 (11.58)	67.00 (15.66)	10.85 (3.10)

Note. For each test, scores below average are highlighted in gray.

Characteristics of the High- and Low-Achievers in this Course

The characteristics of the high-achievers in this course based on their VST and PostT results are shown in Table 3. In this section, students in the 4K level band or with PostT scores above 80 are examined. Firstly, Students A and B had reported at the beginning of the academic year that their learning goal was to attain 700 on the TOEIC. In addition, they both had experience living with native speakers. Student A was lively when working on various classroom tasks, such as pairwork. He was also eager to talk to the instructor in English. Through these talks, the instructor found out that he was living with international students in the same dormitory and studying English on his own to increase his vocabulary. Student B reported that he was born in the US and lived there for five months. When asked about their education prior to university, both students wrote that their schools focused on university entrance examination preparation. Secondly, the rest of the students tended to comment that they wanted to improve their skills or liked this course. Student D was always looking up in class and listened carefully to the instructor. Student I felt that his English skills were not high but mentioned that he would attend every lesson to improve his skills. Student P, the highest PostT scorer in this class, wrote that he wished to work abroad in the future. Student R wrote that he did not like English but wanted to like it.

Other points worth mentioning are that, when compared with the low-achievers (i.e., those in the 2K to 3K level bands with PostT scores below average), the high-achievers were all males and tended to have clearer learning goals, such as a desire to work abroad. Meanwhile, the low-achievers were all females, and they commented that they were learning English for general purposes, such as to enjoy lessons, Disney movies or a TV show *Glee*.

Table 3

Characteristics of the High-Achievers in this CLIL Course

Students	VST (/60)	Post-test (/ 120)	Characteristics
			VST scores 40+ / Post-test scores 80+
A	43	87	a) wants TOEIC 700 / dormitory with English speakers / English test proctoring / often talks to the instructor in L2 / badminton b) focused on tests
B	42	77	a) wants TOEIC 700 / lived in NY (birth – 5 months) / soccer (Barcelona) b) translation / test preparation
D	37	84	a) Singapore for 5 days / always looking up / pitcher / anime b) vocabulary / speaking
I	34	88	a) "I'm not good at English, but I will attend this lectures everytime. And I will improve my English skill in this lecture." / swimming b) translation / textbook sentences
P	30	98	a) wants to work abroad / NZ for a week (junior high) / orchestra b) often: reading, listening / sometimes: writing, speaking
R	29	85	a) doesn't like English but "I want to like English." b) ?

Note. In the far right column, Point *a* shows the student's background information such as learning goals, and Point *b*, what the student's English instructors before university focused on in class.

5.　Discussion

This CLIL course required the participants to use all four skills and handle academic topics/ tasks in the target language. In such lessons, those who had a vocabulary of 4,000 word families at the onset of the course consistently scored above the average scores on all the vocabulary tests. Those who had the vocabulary of 3,700 or 3,400 word families also showed similar performances. Dang and Webb (2014) reported that L1 English speakers would need a vocabulary of 4K (plus proper nouns) to understand 96% of academic spoken texts at university level. Similarly, such learners are said to be able to comprehend 95% of academic texts (Nation, 2006). As the language used in this course was not as high as the one used at universities in an ESL setting, and the teacher adjusted her language to suit the participants' skills, the present members, including those in the 2K to 3K level bands, did not look uncomfortable attending the course. The students were able to continue the interactions with the instructor in discussing the academic topics, but the instructions needed to be repeated several times until the students understood the meanings. When given a turn by the instructor, the students often spent time figuring out with others (using L1) what the instructor had said. Once they were ready to respond, their utterances were simple, consisting of a few words. In addition, the students seemed more satisfied when engaged in daily conversations in class; hence, the instructor made sure she also provided time for them to enjoy less-academic topics. These reactions are understandable as it is easier for them to use the target language more spontaneously in such contexts (cf. Table 1).

In this study, what seems to have reflected more of the participants' three vocabulary test scores in CLIL lessons was their PreT scores than VST scores. This seems a matter of course, as the higher the students had scored on PreT, the more capable they were to understand this CLIL course's vocabulary. (One exception is Students G, who performed unsuccessfully on PreT but successfully on the next two tests.) By doing correlational analyses, the following were found: there was a high correlation between PreT and PostT (r=.86), moderate correlations between PreT and DPostT (r=.69) and PostT and DPostT (r=.79), which were all statistically significant at p < .01. Correlation coefficients of VST and each of the CLIL vocabulary tests were all low. (A correlation of VST and PreT, PostT and DPostT was r=.29, .32 and .48, respectively.) Additionally, it is important to note that the VST used in this study was an original version, which is said to be more suitable for advanced non-native ESL speakers who know more than 5K words (as mentioned in Paul Nation's website https://www. wgtn.ac.nz/lals/resources/paul-nations-resources/vocabulary-tests). Nation also explains that the VST does not measure test-takers' knowledge of multiword units, proper nouns, transparent compounds, marginal words and abbreviations. (The present study was conducted in 2015, and

the studies about the trialing of the latest VST were published around this time.) This will be one of the limitations of the present study; however, since vocabulary experts have used the first version to EFL learners in their studies (e.g., Nakata, 2015), and the participants in this study attended a course similar to the ones native speakers attend in an ESL environment, the original VST was likely to be adequate for the present participants.

The overall tendency of those with a vocabulary of the 3K level was difficult to determine in the present study. Qian and Lin (2020) summarized previous findings concerning the role of the first 3,000 word families and concluded that "word knowledge at the 3,000-word-familiy level has a rather limited role in predicting learners' language proficiency" (p. 77). They mentioned some actual cases of a) EFL undergraduates with a vocabulary of the 3K level not being able to manage required reading at their university, and b) EFL undergraduates' knowledge of the 3K word families weakly correlating with their writing scores or listening comprehension. Perhaps, as the most frequent 3,000 word families are often not used frequently in classroom context (Vilkaitė-Lozdienė & Schmitt, 2020), the students who had had a better command of the high-frequency words and/or PreT words (e.g., the pronunciation) may have performed more successfully in this study.

When the characteristics of high-achievers and low-achievers were compared, what stood out was factors such as their learning goals and gender. The learners who had stated that they hoped to achieve 700 on the TOEIC or work abroad performed successfully in this study. In Japan, several prestigious companies expect their employees to attain TOEIC 700+. For these learners, instrumental motivation may have affected their language learning processes. Students A and B's experience living abroad and/or with native speakers seems to have affected their behaviors and performances as well.[1]

6.　Pedagogical Implications, Limitations and Future Studies

The present study investigated the relationship between learners' vocabulary size and their actual classroom performance, which is something under-researched (Qian & Lin, 2020; Vilkaitė-Lozdienė & Schmitt, 2020). It also looked at Japanese undergraduates' performances in a CLIL course, which is a relatively new type of English learning environment for undergraduates in Japan. Overall, the students participated actively in class, and some informally reported that what they learned in this English course was quite different from what

---

[1]　Additionally, the high-achievers in this class were all males, and low-achievers all females. In Heras and Lasagabaster (2014), it was found that CLIL helped male learners become more interested in foreign language courses. However, there are various findings regarding the relationship between gender and language learners' performances; hence, the present findings should be interpreted with caution.

they had been used to. As the findings showed, however, the students' performances might have improved more if they had had opportunities to communicate with the target language users and had clearer learning goals. In classrooms where the most frequent 3,000 word families are not sufficiently used, being exposed to daily conversations in L2 will help learners deepen their knowledge of the high-frequency words, and therefore support their overall L2 learning processes. In terms of learning goals, there are those with ambitious goals and less ambitious goals. As seen in this study, for the former (the high-achievers in this study), improving their knowledge of mid- and low-frequency words (3,000 and upwards) is useful and will help them greatly in a CLIL course and other environments. For the latter (the low-achievers in this study), if their goal is to engage in a simple conversation and/or watch entertaining movies/ TV dramas in L2, the words used in this CLIL course may be irrelevant and further training in the most frequent 2K to 3K word families is perhaps enough for them (Vilkaitė-Lozdienė & Schmitt, 2020, pp. 90-91). Whether or not undergraduates studying English in Japan should take courses as the present study's should be thought through by both practitioners and students.

There are several limitations of this study. One is that the number of participants was small (n = 20), out of which a) only two had a vocabulary of 4,000 word families, and b) 73% were males. More research is needed to find out if the tendency observed in this study can be generalized to other learners. Another limitation is that this CLIL course was conducted for only three to four months (two 90-minute lessons each week). As previous research found that the effects of CLIL were observed after more than a year (Reynaert, 2019), longitudinal studies are worth conducting to see the true effects of this teaching approach.

7.　Conclusions

The present study compared performances of Japanese EFL university students with a higher initial level of vocabulary and those with a lower one on three vocabulary tests in a CLIL course. The participants (n = 20, CEFR A2 level) were divided into two groups, high and low, according to their Vocabulary Size Test scores carried out before the onset of this course. To examine their vocabulary learning processes, three tests (PreT, PostT and DPostT) were given in class. This semester-long course was taught in the target language, and the students mainly used L2 when interacting with their instructor and working on in-class tasks. The findings showed that students who knew the most frequent 4,000 word families of English scored high on all the tests. However, the performances of the rest of the students, who were in the 2K to 3K level bands, did not reveal straightforward tendencies. For instance, there were those with low VST scores who performed exceptionally well on the tests. As those with VST scores with

4K and above performed successfully on all the tests, it is likely that the students needed 4K and above at the onset of this CLIL course. (The students with a vocabulary of 3,700 word families also performed successfully; hence, the estimated number could be lowered to around 3,700 after observing more results in similar studies in the future.) Additionally, in this study, some characteristics of the high-achievers were as follows: a) had experience interacting with L2 speakers, and b) hoped to attain high scores on the TOEIC, work abroad, improve L2 skills and/or like the target language. Contrarily, the low-achievers tended to comment that they hoped to improve their daily conversation skills. In the future, other variables such as the participants' interactions in class should be investigated as it is highly likely that these affected the students' learning outcomes. In addition, more research is needed to see if similar tendencies will be observed in other CLIL courses.

Acknowledgement and funding

References

Adolphs, S., & Schmitt, N. (2003). Lexical coverage of spoken discourse. *Applied Linguistics*, *24*(4), 425-438.

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

Castellano-Risco, I., Alejo-González, R., & Piquer-Píriz, A. (2020). The development of receptive vocabulary in CLIL vs EFL: Is the learning context the main variable? *System*, *91*, 1-10.

Coyle, D., Hood, P., & Marsh, D. (2010). *CLIL: Content and language integrated learning*. Cambridge: Cambridge University Press.

Coxhead, A., Nation, I. S. P., & Sim, D. (2015). Measuring the vocabulary size of native speakers of English in New Zealand secondary schools. *New Zealand Journal of Educational Studies, 50*(1), 121–135.

Dang, T. N. Y., & Webb, S. (2014). The lexical profile of academic spoken English. *English for Specific Purposes*, *33*(1), 66–76.

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

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

Heras, A., & Lasagabaster, D. (2015). The impact of CLIL on affective factors and vocabulary learning. *Language Teaching Research*, *19*(1), 70–88.

Gierlinger, E., & Wagner, T. (2016). The more the merrier – Revisiting CLIL-based vocabulary growth in

secondary education. *LACLIL, 9*(1), 37-63.

Ikeda, M. (2011). Basic principles of CLIL. In Y. Watanabe, M. Ikeda, & S. Izumi (Eds.), *CLIL (Content and language integrated learning): New challenges in foreign language education at Sophia University, Volume 1: Principles and methodologies* (pp. 1-13). Tokyo: 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 , 122,* 22-28.

Laufer, B. (2020). Lexical coverages, inferencing unknown words and reading comprehension: How are they related? *TESOL QUARTERLY, 54*(4), 1076-1085.

Laufer, B., & Nation, I. S. P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing, 16*(1), 33-51.

Meara, P., & Milton, J. (2003). *X-Lex, the Swansea Levels Test*. Newbury: Express.

Merikivi, R., & Pietilä, P. (2014). Vocabulary in CLIL and in mainstream education. *Journal of Language Teaching and Research, 5*(3), 487-497.

Nakata, T. (2015). Effects of expanding and equal spacing on second language vocabulary learning: Does gradually increasing spacing increase vocabulary learning? *Studies in Second Language Acquisition, 37*, 677-711.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review, 63*, 59-82.

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). Abingdon: Routledge.

Reynaert, B. (2019). CLIL and productive vocabulary acquisition in the Czech context. *Advances in Language and Literacy Studies, 10*(4), 153-160.

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Basingstoke, UK: Palgrave Macmillan.

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

Stæhr, L. S. (2009). Vocabulary knowledge and advanced listening comprehension in English as a foreign language. *Studies in Second Language Acquisition, 31*(4), 577-607.

van Zeeland, H., & Schmitt, N. (2013). Lexical coverage in LI and L2 listening comprehension: The same or different from reading comprehension? *Applied Linguistics, 34*(4), 457-479.

Vilkaitė-Lozdienė, L., & Schmitt, N. (2020). Frequency as a guide for vocabulary usefulness: High-, mid-, and low-frequency words. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 81-96). Abingdon: Routledge.

Xanthou, M. (2011). The impact of CLIL on L2 vocabulary development and content knowledge. *English

Teaching: Practice and Critique, 10, 116-126.

Appendix Lextutor VocabProfile Output for PreT

1K level: ability, average, cause, common, concern, contain, control, development, disagreement, effects, factory, force, forced, gather, industry, object, observe, offer, prevents, proof, proves, senses

2K level: argue, behavior, blame, brain, competition, critics, dirty, dull, exploration, explore, inventor, mechanical, performed, pigeon, prize, programmed, radio, smell, steam, tail, taste, violent

AWL: affected, automatically, brief, constructed, creator, definition, designed, environment, evidence, experts, interaction, issues, labor, nuclear, obtain, outcomes, psychologist, react, require, respond, significant, survey, task, vehicle, vision

OFF LIST: aggressive, Antarctica, BCE, bored, cons, Czechoslovakian, electronic, intellectual, knight, Mars, paraphrase, pros, robotics, silly, vacuum, volcanoes, worldwide

Multiword units: anti-social, hand-eye coordination, get along with, pick up, attention span

---

Note. Multiword units section was added by the author.