

内容言語統合型学習（CLIL）授業における 国内大学英語学習者の語彙学習過程

藤井 里美

Vocabulary Learning Processes of Japanese EFL University Students in a Content and Language Integrated Learning (CLIL) Course

FUJII Satomi

要旨

本稿では、日本で英語を学ぶ日本人大学生が、内容言語統合型学習（CLIL）授業時に気づき、そして、理解した語彙を調査した。被験者（22名）には、各授業後、自身はその授業で学習したと感じた語彙を記してもらい、そのうち、多くの学習者が記入した語彙を後の語彙テストに含めた。また、教員の目標語彙も語彙テスト（プリテスト、ポストテスト、ディレイド・ポストテスト）に含めた。その結果、a) 授業後、目標語彙のうち約半数のスコアが上昇し、b) 学習者が学んだと感じた語彙のスコアよりも目標語彙のスコアの方が高かった。さらに、ポストテストよりもディレイド・ポストテスト時の方が回答の質が向上したケース、また、スコアの低い語彙でも、学習者の誤答を調べると、授業を通じ学習が起きているケースなどが見られた。CLIL授業では、語彙学習が緩やかに起こる可能性があり、テスト作成・採点方法を、さらに調査する必要があると言える。

キーワード：noticing, vocabulary learning/teaching, CLIL, language assessments

Introduction

Content and language integrated learning (CLIL) is “a dual-focused educational approach in which an additional language is used for the learning and teaching of both content *and* language” (Coyle, Hood, & Marsh, 2010, p.1). The term CLIL was adopted in 1994 in Europe (Ikeda, 2015) and, according to Nikula, Dalton-Puffer, Llinares, and Lorenzo (2016), CLIL “has spread especially in Europe since the mid-1990s, and draws on earlier models of bilingual education such as immersion and content-based instruction” (p.1). CLIL students are exposed to abundant comprehensible L2 input which also stimulates their existing knowledge, curiosity and cognitive abilities, and participate in communicative activities while interacting with their teacher and classmates in L2 (Ikeda, 2011, pp.2-3). Now, the CLIL approach has been implemented in other countries, and it is possible to see attempts made in, for instance, Thailand (Chansri & Wasanasomsihi, 2016) and Sri Lanka (Vithanapathirana & Nettikumara, 2020). In 2011, an informative book on CLIL was first published in Japan (Watanabe, Ikeda, & Izumi, 2011), and the present paper reports on how Japanese university students newly introduced to CLIL lessons performed in Japan in 2015.

1. Previous research

According to Schmidt (1990, 1995, 2001), there are three basic levels of awareness, namely *perception*, *noticing* and *understanding*. He asserts that, while we *perceive* everything we are exposed to, we choose what aspect(s) of the input to *notice*, and then *understand* something in the input by, for example, comparing different instances and forming hypotheses. To investigate how language learners’ three levels of awareness operate in a classroom setting, Slimani’s (1989, 1992) research methods are especially worth noting. Slimani, based on Allwright’s (1984) work, asked her participants (university students studying English in Algeria) to recall their *uptaken items*, i.e., items they believed they had learned in class after attending each lesson. Her participants mainly reported on vocabulary items (not grammatical items, which were the main focus of the lessons), and that they showed a tendency to recall items which were not “intended to be taught prior to the lessons” and “arose incidentally in the course of events and became topics in discourse terms” (Slimani, 1992, p.207). Later, other researchers conducted replication studies (some partial replications) and found that students did pay attention to their instructor’s target items (Dobinson, 2006; Fujii, 2008, 2010, 2014; Palmeira, 1995) and that, when some of these target items were later included in the exams, the students’ test performances were successful (Fujii, 2008, 2010; Palmeira, 1995). In Fujii (2014), uptaken words which were not the teacher’s target ones were also tested. The findings showed that the participants were capable of successfully answering the questions, especially when those words had been a) used

in various lesson materials and b) produced frequently in in-class speech. However, one of the main problems of these previous studies is that they did not assess the participants' knowledge of lexical items that had been studied in the previous stages of their learning. There is a likelihood that the students already had some knowledge of these tested words; hence, similar studies using a pre- and post-test design should also be conducted.

Studies on vocabulary learning processes of students in a classroom setting have not sufficiently reported their participants' vocabulary knowledge after the interventions. In studies that measured students' productive and/or receptive vocabulary knowledge in CLIL or content-based courses using a pre- and post-test design, a wide range of different evaluation measures have been applied, including cloze tests, matching exercises (Heras & Lasagabaster, 2015), multiple-choice questions, a general academic vocabulary test adapted from Schmitt, Schmitt, and Clapham's (2001) vocabulary levels test (Li & Cummins, 2019), and/or Laufer and Nation's (1999) vocabulary level tests (Reynaert, 2019). Heras and Lasagabaster (2015) also asked their participants' to write names of given pictures. Even though these previous studies presented interesting research findings, what their participants had noticed and understood in class is unclear, as the previous research did not report the students' actual test answers. Furthermore, in analyzing the collected data, the studies mainly reported the learners' mean scores and other summary statistics for aggregate data. To understand the effects of CLIL or content-based lessons on students' vocabulary learning processes, other ways of investigating and reporting the findings should also be explored, such as interviewing the participants or looking more closely into their disqualified test answers. For instance, by investigating incorrect answers, researchers and practitioners may better understand the degree to which each student misinterpreted or misunderstood the studied term during the learning processes.

2. Research aims

The present study looked at Japanese students' performances in English as a foreign language (EFL) classrooms in Japan using the CLIL approach. The following were investigated:

1. Learners' performances on the instructor's target words on a pre-session test, a post-session test and a delayed post-session test; and
2. Learners' performances on their uptaken words on a post-session test and a delayed post-session test.

3. Research methods

In the present study, a total of 22 students attending a private university in Tokyo participated. Students in an elementary-level freshman course (CEFR, A2 level) were recruited,

whose members majored in Science or Economics. Their total score average of the reading and listening sections on the Test of English for Academic Purposes (TEAP), taken in April 2015, was 88.25 (SD = 0.89) out of 200. Vocabulary knowledge, the focus of the present research, may partly depend on the degree of standardization of students' prior curriculum; however, the present participants' receptive-skill scores of this academic English proficiency test (low-high: 87.00-89.00) may indicate that their levels are similar. The instructor was a Japanese, female, and she used the target language during the lessons. At this university, all freshmen attended an English for Academic Purposes (EAP) course in the first semester, and a "soft CLIL" course in the second semester (in which the lesson aim is geared towards study of language rather than of a specific content or subject, so that several different topics are taught in the target language; cf. Ikeda, 2011, p.10).

The topics covered in the second semester are shown in Table 1 (e.g., 'Trip to Antarctica'), which were selected from a textbook *Contemporary Topics 1* (Solorzano & Frazier, 2009). Using the supplementary materials (a CD and a DVD), the students worked on academic listening and note-taking activities, then speaking and discussion activities about each topic. The course title of the second semester was "People and Technology," as the course aim was to look at the relationship between these two. While working on these different topics, the students were constantly reminded to get ready for their final presentation in December or January in which they needed to talk about some technology of their own choosing (e.g., AI, drones or 3D printers) and how it affects our daily lives. Each student collected newspaper and/or magazine articles about the topic, created Power Point slides, and presented their work twice to two

Table 1
Class Schedule and Research Schedule

Dates	Class schedule	Research schedule
July 2015		- Vocabulary Size Test (July 15 th) - Consent form
October	- History (Trip to Antarctica) - Biology (Genetically modified food) - Original book vs. Movie	- Pre-session test (October 28 th) - Start collecting students' uptaken words - Recording begins
November	- Media (SNS) - Technology (Robotics)	
December	- People and Technology presentations	- Post-session test (December 4 th)
January 2016		- Delayed post-session test (January 22 nd)

different groups of audience. They were also required to submit an essay (the same content as their presentation) afterwards. Additionally, in October, the students gave a presentation with a partner, in which they compared a book they had chosen with its movie version (e.g., *Spirited Away*, *The Eternal Zero*, *Suspect X*).

The main research of this study began in the last month of the first semester and continued until the end of the second semester. In July 2015, the students took the Vocabulary Size Test (Nation & Beglar, 2007; Nation & Gu, 2007) and also received information about the study. The ones who agreed to participate signed the consent form. The Vocabulary Size Test (VST), a test to measure learners' receptive mastery of lexical items, was used, as this study looks at the participants' form-meaning link established at meaning recall level of knowledge and receptive mastery (Schmitt, 2010, pp.79-89) on the pre-, post- and delayed post-session tests. After the summer break, the second semester began from October. First, the students took a pre-session test (hereafter PreT) which contained the instructor's target words ($n = 91$) of the present soft CLIL course. The students were to translate the L2 terms into Japanese. The lessons were audio-recorded from October until January using eight voice-recorders. One recorder was placed on a podium to record mainly the whole-class interactions. The rest were placed on the students' desks to record their interactions. In addition, in November and December, the students wrote down uptaken linguistic items (vocabulary, sentences/phrases, spelling, pronunciation, or others) after each lesson. The first time answering this questionnaire, they wrote recalled items in pencil or black pen. After this process, they were allowed to look at the lesson materials and write additional information (e.g., other items) using colored pens. In December, the students took a post-session test (hereafter PostT), in which included all the words examined in PreT as well as the ones ($n = 29$) the students had claimed in the previous lessons; hence, a total of 120 words were tested. Lastly, in January, the delayed post-session test (hereafter DPostT) was given, taking words from PostT ($n = 22$). Twelve of these terms were those that had been uptaken by multiple students and the rest, the teacher's target words. One of the 12 uptaken words was the teacher's target word and was tested on the three tests. The PostT words were not selected according to how well or poorly the participants had performed on the previous vocabulary tests, as there was not enough time to evaluate their tests by the time DPostT was given. Instead, the selection criteria were based on the degree to which the students had shown interest in each word in class.

The vocabulary tests were given unannounced. (See Appendix for the tested words.) Test items, such as *hand-eye coordination* and *sit still*, may not be suitable for this study; however, since the participants had selected these as learned words, they were included in the tests. In the evaluation process, the correct answer was given one point and the incorrect, zero. The following

were counted as incorrect answers: a) wrong parts of speech, b) wrong translations, and/or c) blanks. For example, if the student's answer was one of the following for the word *confusion*, it was considered incorrect: 困惑する (be confused), 結末 (conclusion), or blanks.

4. Results

The descriptive statistics for the four tests are shown in Table 2. First, their average score on the first to the sixth 1,000-word frequency levels of the VST was 33.73 (SD = 4.29) out of 60. Mean scores of all the other tests showed that the students answered about 50 percent of the words correctly on each test. As can be seen in Table 2, there were two students who did not take one of the vocabulary tests; hence, their scores will be taken off hereafter. Table 3 shows that the total score average of the teacher's target words increased from PreT to PostT. A paired t-test was conducted, and there was a significant difference in the mean scores of PreT and PostT ($t(19) = 4.01, p < .01$), suggesting that PostT score on the target words was higher than PreT score.

Table 4 shows how the learners' performances on the teacher's target words changed after taking the lessons. The performances on 55 out of 91 words (60.44%) improved from PreT to PostT ($t(54) = 8.00, p < .01$). However, on 33 words, the scores declined after the lessons ($t(32) = -7.58, p < .01$). A slight decrease (0.01 to 0.03) in the difference index (*DI*) was observed on 10 words (10.99%), and a decrease of 0.04 or more was seen on 23 words (25.27%). The students' answers on one word were correct on both PreT and PostT, which implies that they all

Table 2
Vocabulary test results

	<i>N</i>	<i>k</i>	<i>M</i>	<i>S</i>	mode	median	midpoint	low-high	range
VST	22	60	33.73	4.29	34.00	34.00	34.00	25.00-43.00	19.00
PreT	21	91	50.57	11.39	57.00	47.00	52.00	32.00-72.00	41.00
PostT	22	120	66.36	15.09	63.00	63.00	69.00	40.00-98.00	59.00
DPostT	21	22	10.62	3.20	6.00	11.00	11.50	6.00-17.00	12.00

Table 3
PreT and PostT results

	<i>N</i>	<i>k</i>	PreT		PostT	
			<i>M</i>	<i>S</i>	<i>M</i>	<i>S</i>
Teacher's target words	20	91	50.25	11.58	55.68	12.74
Students' uptaken words	20	29			10.90	3.25

Table 4**How scores of the teacher's target words changed from PreT to PostT**

	<i>N</i>	<i>k</i>	PreT <i>IF</i>		PostT <i>IF</i>	
			<i>M</i>	<i>S</i>	<i>M</i>	<i>S</i>
Increase in <i>DIs</i>	20	55	0.49	0.27	0.62	0.24
Decrease in <i>DIs</i>	20	33	0.63	0.27	0.57	0.26

Table 5**How *IF* values of the teacher's target words (assessed on all three tests) changed from PreT to DPostT**

	PreT	PostT	DPostT		PreT	PostT	DPostT
<i>concern</i>	0.62	0.77	0.81	<i>dull</i>	0.10	0.45	0.19
<i>aggressive</i>	0.86	0.91	0.90	<i>knight</i>	0.38	0.55	0.62
<i>ability</i>	0.76	0.68	0.81	<i>critics</i>	0.10	0.23	0.05
<i>respond</i>	0.67	0.45	0.52	<i>vision</i>	0.90	0.91	0.90
<i>definition</i>	0.71	0.68	0.76	<i>intellectual</i>	0.57	0.50	0.19
<i>Antarctica</i>	0.71	0.68	0.67	<i>M</i>	0.58	0.62	0.58
				(<i>S</i>)	(0.26)	(0.20)	(0.29)

Note. A paired t-test was conducted, but no significant difference in the mean scores was observed.

knew it perfectly before the main research. There was no case where the scores were zero on both tests.

Table 5 shows the item facility (*IF*) values for the 11 target words examined on all three tests. An *IF* is the percentage of students who correctly answered each item. Of these terms, the performances on six words improved from PreT to PostT (*concern*, *aggressive*, *dull*, *knight*, *critics*, and *vision*); though in two of these cases (*aggressive* and *vision*) the improvement could only be marginal, as most students had already acquired the words before the intervention. The mean scores of the 11 target words tested on PreT, PostT, and DPostT were 0.58, 0.62 and 0.58, respectively.

The students' claimed words were included in PostT as well as DPostT, and the results showed that the mean scores of these words were lower than those of the teacher's target words (Table 6). Eleven terms were tested on the two tests, and the mean scores of PostT and DPostT were 0.38, which is 0.24 and 0.20 lower than that of the teacher's target words, respectively. A closer look at the data showed that the scores of five words improved from PostT to DPostT (*kill time*, *coordination*, *creepy*, *circulation* and *confusion*); however, the scores of four words decreased and two words remained the same. The scores of the word *procrastinate* was always

Table 6
How IF values of the students' claimed words
changed from PostT to DPostT

	PostT	DPostT		PostT	DPostT
<i>eyesight</i>	0.86	0.86	<i>luggage</i>	0.36	0.33
<i>kill time</i>	0.27	0.29	<i>suspect</i>	0.64	0.57
<i>coordination</i>	0.27	0.52	<i>procrastinate</i>	0.00	0.00
<i>creepy</i>	0.05	0.14	<i>sit still</i>	0.32	0.19
<i>intelligence</i>	0.55	0.29	<i>confusion</i>	0.50	0.52
<i>circulation</i>	0.45	0.48	<i>M</i>	0.38	0.38
			<i>(S)</i>	<i>(0.23)</i>	<i>(0.23)</i>

Note. A paired t-test was conducted, but no significant difference in the mean scores was observed.

zero in all the tests.

As for other words whose IF values increased or decreased greatly from PreT to PostT, the following tendencies were observed: a) in the case of the ones whose IF values increased more than 10 percent, the students wrote different parts of speech or L1 counterparts of the words with similar spellings, and tended to write better L1 counterparts on PostT than on PreT; b) in the case of the ones whose IF values decreased more than 10 percent, the students wrote different parts of speech or L1 counterparts of the words with similar spellings, and their PostT performances were not so problematic. (The details are explained in the next section.)

5. Discussion

The students in the present research, who know on average the most frequent 3,000 word families of English, improved their test scores on about a half of the teacher's target words after taking the CLIL lessons. This implies that the instructor's target words do have positive impact on the learners' later performances, as Palmeira (1995) also found in her study. However, the test scores declined in the case of the rest of the target words. The *IF* values of the words listed in Table 7 decreased more than 10 percent from PreT to PostT. For instance, for the word *respond*, the values decreased 0.22, and the students' incorrect answers on PostT were 返信, 反応 or 返事, all of whom its noun forms. (Some of these learners correctly wrote the verb forms on PreT.) There were those who skipped some questions on the tests; for example, in the case of *anti-social*, the students either wrote its noun form 反社会 or left it blank. Furthermore, there were cases where the students performed better on PreT. For *anti-social* and *disagreement*, for instance, all except one had written the correct answers on PreT. Other tendency observed was

Table 7

How IF values of the target words decreased from PreT to PostT and the incorrect PostT answers

	<u>IF</u>		<u>Incorrect answers on PostT</u>
	PreT	PostT	
<i>competition</i>	0.29	0.18	比較 (comparison), まとめ (conclusion), 競争する / きそう (v), [blank]
<i>violent</i>	0.52	0.41	暴力 (n), 暴力する / 暴力をふるう (v)
<i>argue</i>	0.48	0.36	困らせる (annoy), 怒り (anger), 反抗 (rebellion), 賛成する (agree), 酷い (cruel), [blank]
<i>disagreement</i>	0.90	0.77	反対する (v), 反対の / 否定的な (adj), [blank]
<i>anti-social</i>	0.71	0.50	反社会 (n), [blank]
<i>respond</i>	0.67	0.45	返信 / 反応 / 返事 (n)

Note. The translations of the incorrect answers are shown in the parentheses. The rest mean that the students wrote L1 counterparts in the wrong parts of speech: n = noun, v = verb, adj = adjective.

that, for some words, the students showed more eagerness to write answers on PostT. In the case of *argue*, for example, three students wrote the correct answers and the rest did not write anything on PreT, but they wrote various answers on PostT. One possible reason for the decline in scores could have been due to the students' trying to go through 120 questions as quickly as possible and feeling no need to write the precise L1 counterparts. The students' incorrect answers showed that they wrote different parts of speech or L1 counterparts of the words with similar spellings, and their PostT performances were not so problematic.

Contrarily, there were 26 target words whose *IF* values improved more than 10 percent from PreT to PostT, out of which a half of them increased more than 20 percent. The ones that increased more than 30 percent are shown in Table 8. In the case of *constructed*, for example, the incorrect answers on PreT were its present form of the verb 建設する/工事する (construct), 建設中の (under construction), 建設物 (construction), 建築 (building), 保証 (warranty), 成り立つ (consist of) or a blank. On PostT, 13 students wrote the correct answers, of which five were the same students who had written correctly on PreT. One student wrote the correct form on PreT but wrote the present form of the verb 構築する on PostT. Incorrect answers on PostT are shown in Table 8. The only peculiar answer was 保証する (guarantee; the same student wrote this word on both tests); hence, most of the participants roughly or precisely understood the meaning of the word after the lessons. In the case of *dull*, the incorrect answers on PreT were its verb form だらける (be lazy), やつ (fellow), 疑い (suspicion), 休む (take a rest), 掘る (dig), 怠惰な (lazy), 人形 (doll), 汚い (dirty) or a blank. This shows that a few students already had

Table 8

How IF values of the target words improved from PreT to PostT and the incorrect PostT answers

	<i>IF</i>		Incorrect answers on PostT
	PreT	PostT	
<i>constructed</i>	0.29	0.59	構築する / 建設する (v [the present form]), 構成する (compose), 設立する (establish), 構造 (n: structure), 建築物 (n: building), 保証する (guarantee), [blank]
<i>dull</i>	0.10	0.45	やる気のない (lack motivation), 要旨 (main points), だれる (lose interest), 怠惰な (lazy), 味 (taste), 怠ける (be idle), 借金 (debt), [blank], [unreadable handwriting]
<i>designed</i>	0.24	0.73	デザイン (n), デザインする (v), 指定された (be designated), 設計する (be planned, be designed), ~を確認する (confirm), [blank], [unreadable handwriting]
<i>BCE</i>	0.00	0.64	世紀 (century), 国際会議 (international conference), [blank]

Note. See the Note on Table 7.

some knowledge of the term (i.e., だらける and 怠惰な) before the intervention, but a lot of them wrote L1 counterparts that do not seem to have any connection with the tested term. (Note that only one word followed by a line to write an answer was printed on each question. Hence, if the students had written L1 counterparts of *dull* as in, for instance, ‘dull student’ or ‘dull lesson,’ they were both correct.) As can be found in Table 8, on PostT, the learners chose to write better L1 counterparts. (However, the *IF* value declined on DPostT.)

Even though the test performances on the students’ own claimed words were poorer than those on the teacher’s target words, the students’ answers showed some interesting tendencies. First, the following four cases were observed: a) incorrect answers on PostT but became correct on DPostT, b) the parts of speech became correct from PostT to DPostT, c) correct answers on PostT but became incorrect on DPostT, and d) the same incorrect answers on both PostT and DPostT. For instance, a student who wrote まつげ (eyelashes) for *eyesight* on PostT wrote the correct answer 視力 on DPostT. Additionally, the students’ answers for the word *confusion* were verbs such as まざる (be mixed, be confused), なやむ (be troubled) or 困惑する (be puzzled) before but were later correct. (One other student did not write anything on PostT but wrote a verb form 混同する later on.) As such, a few students successfully selected the right parts of speech by the time they took DPostT. Contrarily, there were those who wrote a correct answer before, for example, answering *intelligence* correctly before but later wrote an incorrect answer 知識がある (have knowledge) on DPostT. Lastly, some students continued to write the same

incorrect answers on the two tests; for example, two students wrote 格好 (appearance) for *coordination* on both tests.

Other tendencies seen were that a) there were cases where many of the students' answers were almost correct; and b) the students' L1 answers improved from one test to the next. (Note that this tendency was also observed on the teacher's target words.) An example of the former is the incorrect answers ($n = 16$) of the word *kill time*. Those on PostT were 無駄 (noun: uselessness), 無駄にする (verb: waste), 暇つぶし/ひまつぶし (pastime), 時間がない (have no time), 無駄な時間 (wasted time), 病む (fall ill) or a blank. Incorrect ones on DPostT were 時間切れ (passing the time limit), 暇つぶし/ひまつぶし/時間つぶし (pastime), 暇な時間 (spare time), 時間の無駄 (a waste of time), 傷つける (injure), 殺す (kill) or a blank. These answers imply that many of the students understand some of the features of the expression and that they might be able to reach the correct answer if this expression were, for instance, introduced in context. There were also multiple students who showed the latter tendency. For example, one student wrote the answer of *circulation* correctly in kana as じゅんかん on PostT and later wrote in kanji 循環 on DPostT. For the word *suspect*, three of the students' correct answers had been うたがう (written entirely in kana), 凝う (possibly chosen a wrong kanji) or 予そくする (written using both kana and kanji) previously but later wrote their answers in a more refined manner as 疑う, 疑う, or 予測する, respectively.

The last two tendencies to note are that a) the student who had produced the term in class could not write its correct L1 counterpart on the test; and b) a word claimed by multiple students in colored pen led to an unsuccessful test result. In case of the former, the term *creepy* was claimed by several students on the day their classmate had produced it several times in his presentation. However, only one student answered correctly on PostT, and this student was not the one who had given that presentation. One of the incorrect answers ($n = 21$) of *creepy* was 嫌われている (disliked by others) on PostT and こっそりと (secretly) on DPostT, but the rest seemed that they did not understand the term (e.g., some of the answers were 眠い [sleepy], そうじ [cleaning, tidying] or a blank, and 作物 [crops], ゆっくりとした [without haste] or a blank, respectively). In the case of the latter, the word *procrastinate* was selected by several students, a few of whom incorrectly wrote *procras* in pencil and *procrastinator*, *procrastinater* or *procrastination* in colored pen. The incorrect answers ($n = 22$) of *procrastinate* on PostT were プラスチック (plastic) or a blank. Incorrect ones on DPostT were 宣言する (declare), 実践 (noun: practice) or a blank. These cases show that, even when words are claimed by multiple students, there will be cases where the words are not successfully understood by the learners.

6. Pedagogical implications, Future research and Limitations

The fact that the students' language learning was evidenced in their disqualified test answers and that their performances often improved from PostT to DPostT implies that instructors, evaluators and/or researchers need to carefully consider the role of assessments and how to go about marking students' works in CLIL lessons. An all-or-nothing approach does not seem to be an adequate way to assess the present students' performances, and there need to be more fair and motivating grading systems that effectively cultivate their language skills. For instance, students could take two similar exams assessing the same materials with an interval of a month, compare the two performances and report their own findings to the instructor. It may also be worthwhile to interview each student to find out his or her interlanguage developments. Additionally, as the students wrote answers in more refined Japanese as well as more appropriate parts of speech from one test to the next in this study, future research could take into account both students' L1 and target language developments in CLIL lessons.

The present DPostT only contained 22 items to investigate how well the students remembered the teacher's target words and their own uptaken words. This is a relatively small amount of data, and it is unclear if responses to subsets of 11 items can reflect general patterns of vocabulary development. Hence, test designs need to be improved in the future studies. Additionally, the present data needs to be further analyzed to understand the students' learning processes; for instance, students should be divided into subsets according to their initial level of vocabulary, and levels of improvement should be investigated separately for those subsets.

It should be noted that, in the present paper, the details of the classroom interactions and lesson materials are not sufficiently mentioned. How each word was used in class should be further analyzed to understand why some of the *IF* values of the teacher's target words and the students' words increased or decreased from one test to the next. In terms of lesson materials, the data that especially needs to be investigated in the future is the students' presentation performances. The fact that each student had worked on an individual presentation topic may have contributed to differential vocabulary exposure, and this is likely to have affected the results. In addition, the students' test answers were marked as strictly as possible; hence, the average test scores introduced here are likely to be lower than the ones graded by other practitioners. It is also important to note that there is a possibility of the students' remembering some words tested before, and this might have affected the outcomes. Lastly, the present test answers in L1 are unique to these particular CLIL students, possibly affected by the class members' in-class interactions as well as the contents covered in the lessons.

7. Conclusion

The present study looked at the effects of CLIL courses and the teacher's target words on Japanese EFL university students' noticing and understanding of different lexical items after attending lessons carried out in the target language. Partially replicating Slimani's (1992) research, the following were collected: audio-recordings of the lessons, lesson materials, test scores and questionnaires in which each student ($n = 22$) wrote words he or she believed to have learned in each lesson. The present paper has focused mainly on the students' performances on the VST and three other vocabulary tests (PreT, PostT and DPostT). In the latter three tests, the teacher's target words of the present CLIL lessons were included; in addition, in the two post-session tests, a total of 29 words selected by multiple students were included. The participants were to write the L1 counterpart of each word. The following are the results of the study:

1. PostT scores on approximately a half of the teacher's target words improved compared to PreT scores; and
2. PostT and DPostT performances on the teacher's target words were better than the ones on the students' words.

However, a closer look at the incorrect test answers revealed that the students are in the process of figuring out the precise meanings and/or parts of speech of the words, possibly using the information they had obtained in the lessons. Interestingly, their ways of writing down L1 counterparts improved from one test to the next. These findings imply that CLIL lessons may incrementally develop learners' vocabulary knowledge (Reynaert, 2019), and refine their L2 as well as L1 knowledge. If this is the case in other research sites, practitioners should take into account such tendencies in accurately assessing students' performances.

Acknowledgement

I am grateful to the anonymous reviewer for the invaluable suggestions on an earlier draft of this research project. I would also like to thank the students for participating in this research.

References

- Allwright, R. L. (1984). Why don't learners learn what teachers teach? - The interaction hypothesis. In D. M. Singleton & D. G. Little (Eds.), *Language learning in formal and informal contexts* (pp.3-18). Dublin: IRAAL.
- Chansri, C., & Wasanasomsithi, P. (2016). Implementing CLIL in higher education in Thailand: The extent to which CLIL improves agricultural students' writing ability, agricultural content, and cultural knowledge. *PASAA*, 51, 15-38.
- Coyle, D., Hood, P., & Marsh, D. (2010). *CLIL: Content and language integrated learning*. Cambridge:

Cambridge University Press.

- Dobinson, T. (2006). *Investigating why learners recall certain items of vocabulary from lessons*. Paper presented at 19th Annual EA Education Conference 2006, Australia.
- Fujii, S. (2008). *Factors affecting Japanese EFL high school students' awareness in classroom settings* (Unpublished master's thesis). Sophia University, Tokyo.
- Fujii, S. (2010). In-class input and its impact on Japanese EFL learners' uptake: The case of senior high school students attending a cram school. *ARCLE REVIEW*, 4, 25-37.
- Fujii, S. (2014). *Classroom communication and the vocabulary learning process of EFL students: Lexical items recalled after class and the learners' post performance* (Unpublished doctoral dissertation). Sophia University, Tokyo.
- Heras, A., & Lasagabaster, D. (2015). The impact of CLIL on affective factors and vocabulary learning. *Language Teaching Research*, 19(1), 70-88.
- 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., & Nation, P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing*, 16(1), 33-51.
- Li, J., & Cummins, J. (2019). Effect of using texting on vocabulary instruction for English learners. *Language Learning & Technology*, 23(2), 43-64.
- Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, 31(7), 9-12.
- Nation, I. S. P., & Gu, P. Y. (2007). *Focus on vocabulary*. Sydney: National Centre for English Language Teaching and Research.
- Nikula, T., Dalton-Puffer, C., Llinares, A., & Lorenzo, F. (2016). More than content and language: The complexity of integration in CLIL and bilingual education. In T. Nikula, E. Dafouz, P. Moore & U. Smit (Eds.), *Conceptualising integration in CLIL and multilingual education* (pp.1-25). Bristol: Multilingual Matters.
- Palmeira, K. W. (1995). A study of uptake by learners of Hawaiian. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp.127-161). Honolulu, Hawai'i: University of Hawai'i, Second Language Teaching & Curriculum Center.
- Reynaert, B. (2019). CLIL and productive vocabulary acquisition in the Czech context. *ALLS*, 10(4), 153-160.
- Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 129-158.
- Schmidt, R. (1995). Consciousness and foreign language learning: A tutorial on the role of attention and awareness in learning. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp.1-

- 63). Honolulu, Hawai'i: University of Hawai'i, Second Language Teaching & Curriculum Center.
- Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp.3-32). Cambridge: Cambridge University Press.
- Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Basingstoke, UK: Palgrave Macmillan.
- Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, 18, 55–88.
- Slimani, A. (1989). The role of topicalization in classroom language learning. *System*, 17(2), 223-234.
- Slimani, A. (1992). Evaluation of classroom interaction. In C. J. Alderson & A. Beretta (Eds.), *Evaluating second language education* (pp.197-221). Cambridge: Cambridge University Press.
- Solorzano, H., & Frazier, L. (2009). *Contemporary topics 1: Academic listening and note-taking skill*. New York: Pearson Education.
- Vithanapathirana, M., & Nettikumara, L. (2020). Improving secondary science instruction through content and language integrated learning (CLIL) in Sri Lanka. *International Online Journal of Education and Teaching (IOJET)*, 7(1), 141-148.
- Watanabe, Y., Ikeda, M., & Izumi, S. (2011). *CLIL (Content and language integrated learning): New challenges in foreign language education at Sophia University, Volume 1: Principles and methodologies*. Tokyo: Sophia University Press.

Appendix Words Tested on PreT, PostT and DPostT

	N	Tested words
PreT	91	factory, silly, significant, average, BCE, concern, violent, cause, dull, performed, task, robotics, electronic, intellectual, Czechoslovakian, evidence, prevents, object, senses, programmed, anti-social, gather, creator, aggressive, forced, volcanoes, ability, behavior, effects, proof, bored, pros, interaction, survey, paraphrase, require, taste, explore, pigeon, vision, disagreement, get along with, force, development, common, competition, constructed, Antarctica, affected, steam, exploration, attention spans, outcomes, critics, knight, automatically, smell, worldwide, argue, proves, definition, pick up, respond, psychologist, contain, blame, observe, inventor, mechanical, radio, Mars, cons, vacuum, offer, issues, vehicle, experts, prize, nuclear, brain, industry, labor, tail, obtain, designed, control, environment, dirty, brief, react, hand-eye coordination
PostT	120	sit still, expose, measure, suspect, gambling, Spirited Away, intelligence, sneeze, creepy, kill time, luggage, stiff neck, athletes, folklore, procrastinate, circulation, patent, stereotype, eyesight, dinosaur, germ, confusion, fissure, deprivation, crack, coordination, garbage, bruise, department [+ PreT words]
DPostT	22	intellectual, kill time, concern, eyesight, coordination, creepy, circulation, aggressive, intelligence, luggage, ability, respond, suspect, procrastinate, sit still, confusion, definition, Antarctica, dull, knight, critics, vision