Roles of Vocabulary Knowledge for Success in EMI: Self-Perceptions and Academic Outcomes of Japanese Undergraduates

Takumi Uchihara & Tetsuo Harada

Abstract
This study investigated the relationship between vocabulary knowledge (written and aural receptive vocabulary size and self-rating of vocabulary knowledge) and self-perceptions of four language skills (reading, listening, writing, speaking) targeting undergraduate students in English-medium instruction (EMI) courses in Japan. Their academic performance (course grades and quiz scores) was also compared to their vocabulary knowledge. Results showed that learners with larger aural vocabulary sizes were more confident in spoken language use, and those who self-rated higher on their vocabulary knowledge were more likely to perceive themselves as proficient in productive language skills. Interestingly, learners with larger written vocabulary sizes tended to perceive themselves as less proficient in performing EMI tasks. Results also showed that none of the vocabulary measures were significantly associated with academic outcomes. Interview data suggest that EMI students’ performance could be affected by the complex interplay of various factors, though not necessarily a large vocabulary size alone. Based on these findings, implications are discussed in terms of teaching and assessing vocabulary knowledge in EMI courses.
VOCABULARY KNOWLEDGE AND EMI SUCCESS

Introduction

Internationally, English proficiency has been recognized as a great concern for students and teachers in English-medium instruction (EMI) programs (Macaro, Curle, Pun, An, & Dearden, 2018). Particularly, students’ inadequate second language (L2) proficiency has been considered as one of the greatest barriers to success in EMI programs (Chapple, 2015; Choi, 2013; Floris, 2014; Tatzl, 2011). Research suggests that a main source of this problem stems from students’ lack of vocabulary knowledge (Chang, 2010; Evans & Morrison, 2011; Hellekjær, 2010; Tatzl, 2011). Although it is assumed that vocabulary is a crucial linguistic component for L2 proficiency in general (Nation, 2013), we only have minimal knowledge of the extent to which it is related to students’ successful experience and performance in the context of EMI.

In this paper, several gaps in EMI research are highlighted regarding the relationship between vocabulary and EMI success. Current understanding of this relationship is based on self-reporting from students’ and teachers’ classroom experiences. Except for a few studies (Lin & Morrison, 2010; Harrington & Roche, 2014; Roche & Harrington, 2013), vocabulary knowledge is rarely measured using established vocabulary tests in EMI research. Also, studies focus exclusively on written rather than aural forms of vocabulary knowledge. This study targeted EMI students at the university level in Japan, where English is learned as a foreign language (EFL). Their written and aural vocabulary sizes as well as self-evaluation of vocabulary knowledge were measured and compared to their academic outcomes and self-perceptions of L2 performance spanning reading, listening, writing, and speaking. Based on these findings, we discuss pedagogical implications and provide suggestions for teaching and assessing vocabulary knowledge in order for learners to be successful in EMI contexts.

Proficiency Issues in EMI

The term “EMI” has been used in different ways by different researchers. According to some, EMI focuses on the delivery of content in English without explicit language objectives (see, e.g., Airey, 2016); according to others, it aims for the mastery of content and further development of English skills (Chapple, 2015; Taguchi, 2014). The latter definition of EMI is within the framework of content-based instruction (CBI) models. Brinton and Snow (2017) categorizes EMI as one of the prototypical CBI models—sheltered instruction—where “students who are still developing their L2 are separated from the native speakers for the purpose of content instruction, which is delivered in the students’ L2” (p. 6). Therefore, the authors define EMI as an instructional model which mainly focuses on content but at the same time, consciously or unconsciously, fosters the development of English language skills; the rationale for this choice can be supported by several research findings. In Asian EFL, in some countries where many EMI courses are close to the sheltered-instruction model in the CBI framework (Brinton & Snow, 2017), issues with students’ language proficiency have become more prominent in higher education. For example, in a Chinese context, EMI teachers employ accommodation strategies to compensate for students’ inadequate L2 proficiency, resulting in a limited use of English as the medium of instruction (Jiang, Zhang, & May, 2016). Similarly, in a Korean EMI setting, content teachers express difficulties in keeping to an English-only policy due to students’ low proficiency, and the teachers are often required to assist the students with their English improvement as well as mastery of content (Choi, 2013). In a Japanese EMI setting, despite having studied English for six years before admission, students often lack a sufficient level of L2 proficiency to thrive in an EMI environment (Patrick Ng, 2017). Though the impact of L2 proficiency on content learning has not been well documented (Macaro et al., 2018),
researchers highlight a potential failure to guarantee the quality of content instruction (Floris, 2014; Tatzl, 2011). Content teachers must cater for the clarification of concepts for low proficiency learners, which hinders maintaining the content quality of EMI courses at the same level as when they are delivered in the first language (L1).

Vocabulary knowledge, as a fundamental element of L2 proficiency (Nation, 2013), plays an important role in EMI settings. Data elicited through questionnaire surveys and interviews with students and teachers suggest that insufficient vocabulary knowledge is a major source of learning difficulty, leading to students’ failure or demotivation (Chang, 2010; Evans & Morrison, 2011; Hellekjær, 2010; Tatzl, 2011). Given that vocabulary knowledge is considered crucial for academic success (Nation, 2013), it is reasonable to argue that the larger vocabularies students have, the more confident and successful they will be in performing in EMI. From a pedagogical standpoint, it would be useful for teachers to know how much and what types of vocabulary is needed for students to function well in EMI. Especially, knowledge of words that appear widely across different academic contexts—academic vocabulary—and those that appear frequently in a specific field—technical vocabulary—is crucial for students to be successful in EMI (Hellekjær, 2010; Lin & Morrison, 2010).

From an assessment perspective, testing vocabulary may have a potential to serve screening purposes. In the context of English for Academic Purposes (EAP), vocabulary knowledge of students in EMI foundation programs predicted their performance on university placement and final test scores, confirming the usefulness of vocabulary assessment for screening purposes (Roche, Harrington, Sinha, & Denman, 2016). Though an increasing amount of attention has been directed toward L2 vocabulary acquisition (Webb & Nation, 2017), most previous EMI studies (Chang, 2010; Evans & Morrison, 2011; Hellekjær, 2010; Tatzl, 2011) argue for the central role of vocabulary, based exclusively on students’ self-report data from questionnaires or interviews rather than the measurement of their actual vocabulary size with valid and reliable instruments (for some exceptions, see Lin & Morrison, 2010; Harrington & Roche, 2014).

**Vocabulary Size and Success in EMI**

In order to perform a range of EMI tasks, a solid amount of vocabulary is required. For reading, L2 learners need to master vocabulary sizes of at least 5,000 and ideally 8,000-9,000 word families to understand various genres of authentic materials including academic texts (Schmitt, Cobb, Horst, & Schmitt, 2017). Regarding listening skills, learners need a vocabulary size of 4,000-5,000 word families in order to understand academic spoken English, but 8,000 word families are recommended for sufficient listening comprehension (Dang & Webb, 2014; Schmitt et al., 2017). Similarly, size matters for writing and speaking skills (Laufer & Nation, 1995; Uchihara & Saito, 2016). This brief survey thus suggests that vocabulary sizes of at least 4,000-5,000 word families are necessary for students who aspire to employ the four language skills successfully in EMI courses.

Despite the consensus on the important role of vocabulary knowledge in learning academic content in English, it is rarely measured or explored in relation to learners’ academic success in EMI. One exception is Lin and Morrison’s (2010) study. The researchers measured the receptive and productive vocabulary of EMI university students in Hong Kong and identified the potential difficulties caused by the varying sizes of their vocabulary. Additionally, Harrington and Roche (2014) and Roche and Harrington (2013) investigated the relationship between vocabulary size and success in EMI in higher education. They measured Omani students’ vocabulary size, using yes/no check tests asking test takers to indicate whether they
**VOCABULARY KNOWLEDGE AND EMI SUCCESS**

knew the meaning of words presented individually. The vocabulary scores were compared to their academic performance (i.e., grade point average). Both studies confirm that vocabulary size significantly predicts academic success. Although their findings are revealing, our concern relates to the fact that their findings are solely based on vocabulary measures in written form instead of aural form. Taking account of the modality is important in testing vocabulary arguably because knowing the written form of a word does not necessarily guarantee knowing the aural form of the same word (Milton & Hopkins, 2006). This applies to learners in EFL countries like Japan where their learning is heavily reliant on written input (e.g., reading textbooks). In EMI classes involving not only written activities (e.g., reading textbooks, writing essays) but also spoken activities (e.g., listening to lectures, engaging in class discussion), measuring aural vocabulary size is essential to gain a more comprehensive picture of the relationship between vocabulary size and success.

**Research Questions**

Two research gaps were highlighted above: (a) studies of EMI experience tend to rely on self-reported data and (b) vocabulary knowledge is rarely measured in aural form. This study is designed to respond to these gaps and address the following research questions: to what extent are written and aural vocabulary size measures associated with students’ self-perceptions of L2 performance and academic achievements, and to what extent is students’ self-rating of vocabulary knowledge associated with self-perceptions and academic achievements?

**Method**

This study was conducted in the School of Education’s Department of English Language and Literature at a competitive private university in Tokyo, Japan. Before the beginning of the study, the Department had recently undergone a series of curriculum revisions aimed at increasing EMI offerings in response to societal needs in Japan as well as students’ dissatisfaction with the traditional curriculum (Harada, 2017). Consequently, the Department started to offer EMI for upper-division content courses and two introductory “bridging” content-based instruction courses in 2016 whereby freshmen were to develop EAP skills so that they would succeed in EMI content courses. Our research focus was on the upper-division EMI courses as a whole and on one EMI class in particular; the primary goal of this class was to develop students’ subject knowledge rather than improve English language skills, though further development of English skills was considered to be a by-product of taking EMI courses.

**Design**

We employed an explanatory sequential mixed-methods design (Kim, 2013). In this approach, quantitative analysis precedes qualitative data collection so that quantitative findings can be explained by subsequent analyses based on the qualitative data. First, we conducted quantitative data analysis consisting of two parts: (a) comparing three vocabulary measures and perceptions of language use in the EMI courses collected from 35 undergraduate students, and (b) comparing these vocabulary measures with academic achievement data collected from 16 out of the 35 students who took one of the EMI courses taught by one of the authors. Next, we conducted follow-up semi-structured retrospective interviews with 10 out of the 16 students who took the same EMI course.

**Participants**

Thirty-five junior and senior Japanese-speaking students, including one non-credit student (\(M_{\text{age}} = 21.8\), 18 females, 17 males), participated in the study. Their English proficiency ranged from an intermediate to advanced level according to the scores on the Test of English for International Communication (TOEIC), comprising reading and listening sections (\(M = 723.9\), ...)
VOCABULARY KNOWLEDGE AND EMI SUCCESS

$SD = 114.5$, Range $= 480-935$) and the paper-based Test of English as a Foreign Language (TOEFL) ($M = 538.8$, $SD = 38.0$, Range $= 473-637$). Four participants’ length of residence in English-speaking countries exceeded four years before matriculation. Though they were not typical EFL learners, we included their data to reflect the reality of the increasing number of EMI courses in Japanese universities featuring a heterogeneous population including returnees as well as domestic students (Taguchi, 2014). All participants had attended at least one EMI course at the time of data collection.

**Vocabulary Measures**

Two objective measures (written and aural vocabulary size tests) and one subjective measure (self-rating vocabulary scale) were adopted to assess participants’ vocabulary knowledge. In general, self-perceived L2 proficiency is reported to be moderately correlated with actual L2 proficiency scores, but the correlations are far from perfect (Ross, 1998). Therefore, the two types of measures, subjective and objective, are not identical but independent of each other, suggesting that the combination of the two provides a finer indication of overall vocabulary knowledge.

**Vocabulary Levels Test (VLT).** The VLT is a receptive vocabulary test on a word-matching task (see Appendix 1) measuring meaning recognition of word knowledge in written form (Schmitt, Schmitt, & Clapham, 2001). The test contains 30 items in each section of four word frequency levels (2,000, 3,000, 5,000 and 10,000) and academic vocabulary from the Academic Word List (AWL) (Coxhead, 2000). In this study, we used Version 2 of the test available in Appendix 1 from Schmitt et al. (2001, pp. 82-88). The participants completed one practice set of the task (a block of three items) before taking the VLT to familiarize themselves with the task format. They spent approximately 20 minutes ($M = 19.1$) completing the task.

**Listening Vocabulary Levels Test (LVLT).** The LVLT is a newly developed measure of aural receptive vocabulary size (McLean, Kramer, & Beglar, 2015). The test assesses meaning recognition through a multiple-choice task (see Appendix 2). It taps into “on-line phonological processing of lexical forms” so as to mirror “the demands made on a listener’s aural vocabulary knowledge in most authentic L2 listening situations” (McLean et al., 2015, pp. 742-744). The test contains 24 items in each section of five word frequency levels (1,000 to 5,000) and 30 items in an academic vocabulary section (AWL). Examinees are first aurally presented with a target item (e.g., see) and then a short defining context sentence with the target item embedded (e.g., They saw it) only once. There is a 5-second pause during which examinees answer after the aural presentation of each item. It takes 30 minutes to complete all sections. In this study, the participants first went through practice questions (two items) and then completed the LVLT in which the items and questions had been audio recorded by a male native speaker of General American English, the variety of English primarily taught in Japan (audio files and task materials are available at https://wilmina.academia.edu/StuartMcLean/Instruments).

**Self-rated vocabulary.** In addition to two receptive vocabulary measures, we used self-rating of vocabulary knowledge to assess students’ perceived lexical proficiency. The participants indicated how confident they were in their vocabulary knowledge on a 6-point Likert scale ($1 = not confident, 6 = very confident$).

**Self-perceptions of performance across Four Skills**

Building on the earlier versions of questionnaires used to measure students’ perceptions of language skills (Berman & Cheng, 2001; Evans & Morrison, 2011), we developed a questionnaire consisting of four parts, each of which represented listening, speaking, reading and writing skills. After iterative revisions and modifications through consultation with an
VOCABULARY KNOWLEDGE AND EMI SUCCESS

experienced EMI instructor, the final Japanese version of the questionnaire comprised six items in the listening section, ten items in the speaking section, eight items in the reading section and five items in the writing section (see Appendix 3 for an English version of the questionnaire). The participants answered each question on a 6-point Likert scale (1 = strongly disagree, 6 = strongly agree).

**Academic Achievement**

Course grades and quiz scores were collected from a sub-set of the participants (n = 16) who took the same EMI course (the overview of this EMI context is provided below). A weekly quiz was created based on a course textbook and given every class, which required students to (a) explain a few concepts and (b) answer an open-ended question (Appendix 4). Although the sample size is small, a great number of dropouts reflected the great challenge faced by Japanese EMI students; approximately 40 students were initially enrolled in this course, about 10 (25 to 30%) of whom dropped it during the first or second week after realizing that it required an advanced level of English and many demanding assignments. Further, several more students withdrew from the class later on, resulting in 21 students remaining in total, 16 of whom participated in the study. Despite this limitation, a close investigation into such an intact class is ecologically valid, giving deeper insights into what shapes a successful EMI class in reality.

**Target EMI context.** The focal EMI course was on English language teaching and titled *Content and Language Integrated Learning (CLIL).* The Department of English Language and Literature offered this course as an elective in which the students learned the concept of CLIL and its application to language teaching practice. The instructor was a Japanese native speaker with 16 years of EMI teaching experience and an expert in the target subject. Twenty-one students were enrolled in the course and 16 of them participated in the study. The class met for 15 weeks (once a week) and every lesson lasted 90 minutes. The course structure was built on five major components: (a) reading assignments prior to class, (b) a weekly quiz, (c) a lecture from the instructor, (d) students’ presentations, and (e) group discussions during the lecture and students’ presentations. Before the class, students were expected to read two course textbooks written in English (approximately 15 pages per week). At the outset of the class, they took a weekly quiz (15 minutes). The subsequent lecture delivered by the instructor lasted approximately 30 minutes; he first reviewed the content taught the previous week and then introduced new topics using PowerPoint slides. Subsequently, approximately 20 minutes were set aside for each of the two students to give a presentation based on the assigned chapters. The presenters led more than one discussion question so as to increase opportunities for the other students to discuss the content in groups. The students were divided into 6 groups of four or five including one teaching assistant (TA). Since there were four Japanese-speaking TAs, who were all graduate students majoring in applied linguistics, each group normally had one TA as a facilitator.

**Course grades and weekly quiz scores.** Weekly quiz scores were collected from the 16 students (8 females, 8 males) who were enrolled in the focal EMI course for one semester. The scores were averaged to yield a single score per participant. Course grades were collected from the same participants and quantified by assigning numbers as follows: A+ (90-100%) = 5, A (80-89%) = 4, B (70-79%) = 3, C (60-69%) = 2, F (1-59%) = 1. The breakdown of the final grades is presented below (see Appendix 5 for a detailed description):

- Attendance: 5%
- Weekly quizzes: 25%
VOCABULARY KNOWLEDGE AND EMI SUCCESS

• Presentation/participation in class discussion 10%
• Teaching portfolio 60%

Interviews

Based on descriptive and inferential analyses on three vocabulary measures, four-skill self-perceptions scores and academic achievements, we formulated several interview questions related to three broadly predefined themes: (a) general impression or experiences of the EMI course, (b) difficulties with vocabulary use and four skills required in specific situations, and (c) vocabulary learning experience in general. Out of the 16 students who had attended the focal EMI course, 10 (5 females, 5 males) volunteered to participate in the interview. The interviews were recorded with a digital voice recorder, transcribed by the first author and analyzed following Lune and Berg’s (2017) qualitative analysis.

Procedure

The tests were conducted individually in one session per participant with one investigator (one of the authors). Each individual meeting was arranged toward the end of the semester (Weeks 13, 14 and 15) or after the semester was over. The questionnaire about self-perceptions was sent to each participant prior to the experiment and they were asked to bring the completed questionnaire or e-mail it to the investigator before the testing session. In a quiet room, the participants first completed a language background sheet. After the investigator checked whether all the questions were clear to the participants and answered without any omission, they took three vocabulary tests (the self-rating scale was embedded in the language background questionnaire and had been already completed). The two vocabulary tests were administered under the supervision of the investigator with a five-minute break allowed between each test. They first took LVLT wearing a headset and then VLT. Since the vocabulary tests were administered in a laboratory setting, confounding effects observed when administered in the classroom environment were mostly avoided (e.g., consulting with other students). Approximately five months after the testing sessions, ten students from the focal EMI course volunteered to participate in semi-structured interviews with the investigator. Each interview was conducted in Japanese either on or off campus.

Results and Discussion

In this section, results of quantitative and qualitative analyses on vocabulary knowledge and EMI success (self-perceptions and academic performance) are provided along with interpretations and discussions based on the findings.

Vocabulary Measures

Table 1 presents the descriptive statistics of VLT, LVLT and self-rating scores. VLT scores were initially tallied by summing all the sub-scores from each section (2,000, 3,000, 5,000, 10,000 and AWL). According to a Cronbach’s alpha analysis, the reliability of the VLT scores initially appeared low (α = .67). However, the reliability value was higher (α = .70) if the 2,000-level section was excluded (α = .70–.80, an accepted level; Larson-Hall, 2010). Thus, we calculated the sum of the four sections as final scores. Similarly, LVLT scores were calculated as the sum of the five sections (2,000, 3,000, 4,000, 5,000 and AWL) (α = .77) exclusive of the 1,000-level section (α = .73). In the 10,000-level section (VLT), we eliminated one outlier who scored 18 points (>3 SDs) for further statistical analysis. This participant was an advanced learner (TOEIC scores = 850) and self-reported rich learning experiences in English-taught programs in higher education (2.5 years). According to the criteria for mastery of each frequency and academic vocabulary section (87% or above; Schmitt et al., 2001), the majority of students
mastered the 2,000 (M = 97%) and 3,000 (M = 88%) levels and academic words (M = 87%) in written form and the 1,000 level (M = 97%) in aural form, though failing to master the aural form of the most frequent 2,000 word families and above. Specifically, 54% of all participating students (19/35) and 69% of the participants enrolled in our focal EMI course (11/16) failed to achieve the 2,000 level in aural form. Similarly, 11% (4/35) and 13% (2/16) failed to master academic vocabulary in aural form. These findings confirm a discrepancy between written and aural vocabulary knowledge, which could impose a significant burden on students’ understanding of lectures. Results of the interrelationship among the measures showed that the VLT was strongly correlated with the LVLT (r = .73, p < .01). Self-rated vocabulary significantly correlated with the LVLT (r = .37, p < .05) but not with the VLT (p > .05), suggesting students’ perception of vocabulary knowledge is more closely related to aural vocabulary rather than written vocabulary.

Table 1
Descriptive Statistics of Vocabulary Measures: VLT, LVLT, and Self-Rating Scores

<table>
<thead>
<tr>
<th></th>
<th>MPS</th>
<th>M (%)</th>
<th>SD</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>VLT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2,000</td>
<td>30</td>
<td>29.1 (97)</td>
<td>0.9</td>
<td>26-30</td>
</tr>
<tr>
<td>3,000</td>
<td>30</td>
<td>26.5 (88)</td>
<td>2.6</td>
<td>19-30</td>
</tr>
<tr>
<td>5,000</td>
<td>30</td>
<td>18.7 (62)</td>
<td>3.7</td>
<td>10-29</td>
</tr>
<tr>
<td>10,000</td>
<td>30</td>
<td>3.1 (10)</td>
<td>1.7</td>
<td>1-8</td>
</tr>
<tr>
<td>AWL</td>
<td>30</td>
<td>26.2 (87)</td>
<td>2.6</td>
<td>21-30</td>
</tr>
<tr>
<td>Total</td>
<td>120</td>
<td>74.9 (62)</td>
<td>8.8</td>
<td>55-105</td>
</tr>
<tr>
<td>LVLT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1,000</td>
<td>24</td>
<td>23.3 (97)</td>
<td>0.7</td>
<td>22-24</td>
</tr>
<tr>
<td>2,000</td>
<td>24</td>
<td>20.2 (84)</td>
<td>2.0</td>
<td>16-24</td>
</tr>
<tr>
<td>3,000</td>
<td>24</td>
<td>16.7 (70)</td>
<td>2.9</td>
<td>11-22</td>
</tr>
<tr>
<td>4,000</td>
<td>24</td>
<td>18.0 (75)</td>
<td>2.6</td>
<td>13-23</td>
</tr>
<tr>
<td>5,000</td>
<td>24</td>
<td>13.5 (56)</td>
<td>2.5</td>
<td>8-21</td>
</tr>
<tr>
<td>AWL</td>
<td>30</td>
<td>22.2 (74)</td>
<td>2.9</td>
<td>14-27</td>
</tr>
<tr>
<td>Total</td>
<td>126</td>
<td>90.7 (72)</td>
<td>9.3</td>
<td>69-115</td>
</tr>
<tr>
<td>Self-rating</td>
<td>6</td>
<td>3.2</td>
<td>0.8</td>
<td>2-5</td>
</tr>
</tbody>
</table>

N = 35 except for F10 (N = 34). VLT = Vocabulary Levels Test; LVLT = Listening Vocabulary Levels Test; MPS = maximum possible score; AWL = academic word list.

Self-perceptions of Language Use

For students’ self-perceptions of L2 performance, all items were consistent (α = .97) and so was each sub-category: listening (α = .95), speaking (α = .94), reading (α = .90) and writing (α = .90). Assuming that 3.5 is the neutral value on a 6-point scale, most students positively perceived their L2 performance (overall scores, M = 3.9, SD = 0.8), while some had low perceptions (minimum scores = 2.5). Sub-category data reveal that students were more confident in receptive language skills (M = 4.6, SD = 1.1 for listening, M = 4.1, SD = 0.8 for reading) than in productive language skills (M = 3.5, SD = 1.0 for speaking, M = 3.8, SD = 1.0 for writing), among which the mean of speaking was the lowest (see Appendix 6 for descriptive statistics of sub-category items).

Academic Achievements
Regarding the course grades of students enrolled in the EMI course, most of them passed the course with good standing ($n = 16, M = 3.8, SD = 1.1, \text{Range} = 1-5$): the number of students who gained A$^* = 4$, A = 7, B = 3, C = 1 and F = 1. They also earned relatively high scores on weekly quizzes ($n = 16, M = 17.0 \text{ out of } 20.0, SD = 1.4, \text{Range} = 13.9-19.5$).

**Relationship between Vocabulary Knowledge and Self-Perceptions**

To examine the relationship between vocabulary knowledge and self-perceptions of the four skills, correlation analyses were conducted between three vocabulary measures (VLT, LVLT, self-rating) and self-perception scores. Spearman rho correlation was used due to non-normality of the self-perception scores. Table 2 shows that the LVLT significantly correlated with the average speaking scores ($r_s = .34$), while neither the VLT nor the self-rated vocabulary showed significant correlations with the self-perception scores. However, a further examination of each sub-category item (see Appendix 7 for a summary of correlations) revealed that the VLT was significantly and negatively correlated with one speaking item—*engaging in classroom discussion actively*, $r_s = -.37$. Additionally, significant correlations were found between self-rated vocabulary and self-perceptions concerning productive language use (speaking: $r_s = .36-.50$ and writing: $r_s = .42$ and .40). This indicates that students with confidence in their vocabulary knowledge likely view themselves as proficient speakers and writers. Overall, results suggest that students’ lack of aural vocabulary may be a major factor resulting in unsatisfactory EMI performance especially when production skills are required, confirming earlier research findings (Chang, 2010; Evans & Morrison, 2011; Hellekjær, 2010; Tatzl, 2011). However, concerning the relationship between written vocabulary size and four-skill perceptions, the results countered our expectation in that we found a negative association between VLT and a speaking-related perception score.\(^3\)

A possible reason for this negative correlation may be attributed to the differences in the construct that VLT and LVLT purport to measure. LVLT is designed to mirror the transitory and cognitively taxing nature of spoken language by means of involving task immediacy as examinees are only allowed to listen to each item once and are given five seconds to answer. Conversely, in VLT, they can spend as much time as they want and read each target item and question several times, thereby lowering the degree of task immediacy. With these task differences in mind, we speculate that VLT taps intentionally or consciously learned vocabulary (explicit knowledge), whereas LVLT reflects unintentional or nonconscious use of previously acquired vocabulary (implicit knowledge) (Hulstijn, 2002). Since time pressure is recognized as a crucial factor in measuring implicit knowledge (Ellis, 2005), LVLT likely taps linguistic knowledge learners are not aware of, whereas VLT largely elicits learners’ consciously accumulated knowledge of L1 meaning-L2 form association through its decontextualized multiple-matching task. Given that EFL learners tend to rely on explicit language-focused learning with word lists and flashcards (Nation, 2013; Webb & Nation, 2017) just as were our participants, their learning experience may be predominantly predicated on conscious accumulation of explicit knowledge rather than unconscious input of implicit knowledge (Hulstijn, 2002). Consequently, students who spend more time and effort learning L2 vocabulary explicitly may face a discrepancy between the expectations built upon their personal learning experience and the realities of actual language use in classroom contexts, which may eventually lead to disappointment and lower self-perception (Horwitz, 1988). Therefore, we speculated that the negative correlations were attributed to the gap between what they had in terms of explicit knowledge and what they wanted to achieve by drawing on implicit knowledge.
Our speculation was tentatively supported by the following post hoc analysis. We excluded the data from four participants with greater length of residence in English-speaking countries (> four years) and ran another correlation analysis between VLT scores and self-perceptions. We did this in order to examine whether the negative correlation would become more salient when our data only represented the learners whose experience predominantly relied on explicit learning exclusive of the four proficient learners with greater implicit learning experience. The resulting correlation between VLT scores and overall self-perceptions marked an increase in its effect size and reached statistical significance: \( r_s = -.22, p > .05 \rightarrow r_s = -.47, p < .025 \) (Bonferroni corrected). This finding may support our speculative account of negative correlations between written vocabulary size and self-perceptions, yet it remains tentative as our measures were not designed to purely assess explicit or implicit knowledge.

Table 2

| Correlations between Vocabulary Measures and Self-Perceptions |
|-------------------|-------------------|-------------------|
|                   | VLT | LVLT | Self-rating |
| Overall scores    | -.22| .26  | .29          |
| Listening         | -.22| .18  | .17          |
| Speaking          | -.09| .34* | .33          |
| Reading           | -.24| .20  | .25          |
| Writing           | -.22| .25  | .32          |

* indicates \( p < .05; N = 35 \). VLT = Vocabulary Levels Test; LVLT = Listening Vocabulary Levels Test.

To further probe into the relationship, another set of correlation analyses were performed between self-perceptions and each frequency and academic vocabulary (AWL) section of VLT and LVLT. VLT scores did not correlate with averaged scores for each skill (see Appendix 8 for this and the correlations between each section and sub-category items). Regarding the LVLT (Table 3), the 4,000-frequency level was significantly correlated with the listening average scores, and the 5,000-frequency level was significantly correlated with the listening, speaking, and writing average scores (see Appendix 9 for the correlations between each section and sub-category items). These findings suggest that with a vocabulary of 4,000 and 5,000 word families, students may be more likely to perceive themselves as competent listeners and speakers in EMI situations.

Table 3

| Correlations between Frequency and Academic Vocabulary (LVLT) and Self-Perceptions |
|-------------------|-------------------|-------------------|-------------------|-------------------|
|                   | F2 | F3 | F4 | F5 | AWL |
| Overall scores    | .22| -.08| .26| .45*| .09 |
| Listening         | .26| .03 | .35*| .37*| .04 |
| Speaking          | .22| .15 | .33| .49*| .15 |
| Reading           | .16| .13 | .11| .26 | .10 |
| Writing           | .10| .14 | .21| .37*| .10 |

* indicates \( p < .05; ** \) indicates \( p < .01 \). N = 35. LVLT = Listening Vocabulary Levels Test; F2 = 2,000 frequency level; F3 = 3,000 frequency level; F4 = 4,000 frequency level; F5 = 5,000 frequency level; AWL = academic word list.
Relationship between Vocabulary Knowledge and Academic Achievements

Correlation analysis was conducted on three vocabulary measures and academic achievements (course grades and quiz scores). There was no statistically significant correlation between the vocabulary measures and grades or quiz scores, and further analysis with frequency and academic vocabulary sections did not reveal any other significant correlations. The absence of correlations may lie in the fact that the grades were based on multiple grading components including attendance, weekly quizzes, presentations, participation in group and/or classroom discussion, and the design of a teaching portfolio. Therefore, finding a clear relationship between such composite scores and vocabulary knowledge may not be feasible. Another reason might relate to a mismatch between the type of knowledge measured with the vocabulary tests and what the EMI tasks assessed. Whereas the VLT and LVLT assessed students’ receptive knowledge, the EMI class required their productive knowledge as well as content knowledge (e.g., quizzes, oral presentations); therefore, subject-specific and productive vocabulary may have been more relevant.

Interviews

Interview analysis was incorporated into this study with the aim of gaining a deeper understanding of unexpected quantitative findings: (a) negative correlations between VLT scores and self-perceptions of language use and (b) the absence of significant correlations between vocabulary measures and academic achievement. In general, the interview data of students’ learning experiences in EMI indicate perceived difficulties related to listening (e.g., understanding lecturers), speaking (e.g., using right words in discussion), reading (e.g., explaining technical terms in the quiz). Notably, difficulties dealing with technical terms were frequently observed across the four skills, which agrees with previous EMI studies (Evans & Morrison, 2011; Hellekjær, 2010; Tatzl, 2011). The analysis also indicates factors related to strategies and resources—communication strategies, code-switching, asking TAs for help, checking a glossary, studying PowerPoint slides beforehand—to cope with difficulties students faced.

The interview data highlight complex relationships between vocabulary knowledge and academic achievement (Table 4). Achieving an A+ seems feasible for students who mastered both written and aural academic words (S1, S3), and they rarely experienced difficulties in the class. Three students, S6, S7 and S8, who earned an A but lacked the basic aural vocabulary of the 2,000 level experienced difficulties understanding lectures and students’ presentations. Consequently, the lack of understanding of content might have made their quiz scores (16.9, 15.8, 15.7) lower than the average (M = 17.0). Notably, S10 who only mastered basic vocabulary levels (2,000 written and 1,000 aural knowledge) obtained an A just as did the other three students whose vocabulary knowledge was beyond the basic level. One possible explanation is that S10 capitalized on the resources available to compensate for his lack of L2 proficiency. He mentioned “I had a hard time doing all the assigned readings before class, so I studied the PowerPoint slides posted by the instructor to understand the content” (our translation). Another case worth further exploration is S9, who had a relatively larger vocabulary and gained quiz scores slightly higher than the average (17.3) but earned the lower grade, C. The low grade was attributed to a failure to submit several assignments instead of proficiency-related insufficiency. S2, S4 and S5 lacked vocabulary knowledge, and S4, who gained low scores on weekly quizzes (13.9), mentioned in the interview, “one of the hardest things in the course was reading assigned chapters every week because so many technical terms and difficult words appeared in the
textbooks.” S2 and S5 with small vocabulary sizes nevertheless gained high quiz scores (18.6 and 18.2), but earned a B because they failed to attend the class and/or submit assignments.

The analysis highlights the important roles of strategy use and resources available for students who generally fell short of the minimum level of vocabulary sizes (i.e., 4,000-5,000). In particular, a considerable variability in basic aural vocabulary knowledge across individuals (11 out of 16 did not master the 2,000 level) indicates that some students might have greater difficulties understanding the lectures (Dang & Webb, 2014). Interestingly, however, most of them (15 out of 16) passed the course. Based on this fact and the interview data, some alternative ways appear to make up for the lack of vocabulary knowledge in order for the students to function well in EMI settings. When unable to come up with the right words in group discussion, some students made up for them using communication strategies such as paraphrasing or defining the words they did not know how to say in English, listing synonyms or near-synonyms to have others infer their message, or explaining them with concrete examples. In addition, students remarked in the interviews that they relied on a variety of useful resources available in the course. They benefited from human resources, or the presence of a TA in each group as S7 mentioned “when I didn’t know how to express myself, I just tried saying relevant words randomly and asked the TA ‘Am I understood?’ and then he paraphrased what I meant for me.” Another relates to linguistic resources, or their L1 (Japanese). Since the instructor, TAs, and students all shared the same L1, some interviewees benefited, like S8: “when having difficulty expressing my opinions in English, I said Japanese words instead, and then others in my group helped me find out how to put them in English.” Learning resources were also helpful for struggling students. S4 appreciated the glossary of technical terms at the back of the textbook in preparation for the weekly quizzes. These findings and previous EMI literature (Björkman, 2011) suggest that the students took full advantage of strategies and resources available in the EMI course as a way to compensate for the lack of L2 vocabulary knowledge. This finding may be another account for the lack of significant correlations between vocabulary knowledge and academic achievement.

Table 4
Summary of Vocabulary Size, Academic Achievement, and Perceived Difficulties in an EMI course of Ten Undergraduate Students

<table>
<thead>
<tr>
<th>VLT</th>
<th>LVLT</th>
<th>Academic</th>
<th>Interview (difficulties in EMI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>F3</td>
<td>AWL</td>
<td>F2</td>
<td>F3</td>
</tr>
<tr>
<td>S1</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>S2</td>
<td>✔</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>S3</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>S4</td>
<td>✔</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>S5</td>
<td>✔</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S6</td>
<td>✔</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S7</td>
<td>✔</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S8</td>
<td>✔</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>S9</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>S10</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
</tbody>
</table>

Note. QS = quiz scores, CG = course grades. All students mastered VLT 2,000 level and LVLT 1,000 level; none of them mastered the levels above VLT 3,000 or LVLT 2,000. The criteria of mastery is based on Schmitt et al.’s (2001) suggestion (26/30 words, or 87%). UR = using right words in group discussion, UT = understanding technical terms, FR = finishing reading before
the next class, MO = making oral presentations, UC = understanding the content of reading material, UL = understanding lectures and/or presentations, SL = speaking within a limited time during discussion, ET = explaining technical terms in quiz, CD = coping with difficult words in the textbook.

**Vocabulary as a Construct in EMI**

Findings based on the interview data and correlations between vocabulary measures provide important insights into students’ perceptions of vocabulary knowledge. While self-rated vocabulary was not associated with VLT scores, there was a significant correlation with LVLT scores. These findings suggest that students’ perception of vocabulary knowledge may not simply build on the knowledge of form-meaning connections elicited through the written receptive vocabulary measure. Alternatively, it may be largely related to the ability to use L2 words implicitly or fluently in spoken contexts. Further, we found that students’ perceived vocabulary knowledge was exclusively associated with their self-perceptions pertinent to productive language use (speaking and writing). This implies that learners’ perceived lexical knowledge may be substantially affected by noticing the gap in their knowledge when producing the L2 and encountering linguistic problems (Swain & Lapkin, 1995). This argument was supported by the interview data from 7 out of the 10 interviewees who pointed out their inadequate vocabulary knowledge surfaced in producing the L2, particularly when required to speak under such time pressure as in group discussion and oral presentations. Taken together, the construct of learners’ perception of vocabulary knowledge is complex and appears to encompass vocabulary knowledge called upon particularly in performing L2 production, which involves a higher processing load.

**Conclusion**

In sum, quantitative analysis reveals that (a) aural receptive vocabulary size (especially, 4,000-5,000 levels) was more notably associated with self-perceptions of spoken language use, (b) written receptive vocabulary size was negatively related with self-perceptions, (c) self-rated vocabulary was specifically associated with self-perceptions of productive language use, and (d) there was no significant correlation between vocabulary knowledge and academic performance. Subsequent qualitative analysis indicates that students experienced difficulties across the four skills especially when using appropriate words in group discussion and understanding technical terms. It also indicates that the frequent use of strategies and resources available in the course helped them complete the course successfully.

We acknowledge several limitations of this study. First, we should be cautious about the generalizability of the findings based on the participants majoring in English Language and Literature at a competitive university. Notably, significant relationships for academic vocabulary were not found in the study. However, this finding does not diminish the value of learning academic words, given that most of the participants had already mastered academic vocabulary ($M_{AWL} = 87\%$). This notwithstanding, we should note that although considered to be advanced learners in general, our participants showed great variability in aural vocabulary knowledge. Second, given that the vocabulary self-rating measure comprised one Likert-scale item, its reliability may not be sufficiently robust. To increase the reliability, it would be necessary to include more than one item to reduce measurement errors. Third, measuring productive vocabulary knowledge will provide deeper insights into the relationship between vocabulary knowledge and EMI performance. Although this study only adopted receptive measures, given the nature of the EMI course inevitably involving productive language use (e.g., oral
presentation, essay writing), our understanding of the role of vocabulary knowledge in EMI will benefit from integrating productive vocabulary measures such as free or controlled production tests (Lin & Morrison, 2010). Fourth, a 5-month gap between the interviews and the end of the course might not have allowed the interviewees to recall their experience as clearly as we expected. Besides the sequential mixed-methods design we adopted, future investigation using a concurrent design (Kim, 2013) may be warranted. Finally, future research should explore the relationship between technical vocabulary knowledge and EMI success. A corpus-based study (Durrant, 2014) illuminates the importance of “discipline-specific” vocabulary knowledge, questioning the one-size-fits-all usefulness of “generic” academic vocabulary (e.g., AWL). The importance of technical vocabulary is confirmed by previous EMI studies (Evans & Morrison, 2011; Hellekjær, 2010; Tatzl, 2011; see also Basturkmen, this issue) and the current study as 7 out of the 10 interviewees expressed difficulties with technical vocabulary.

Implications and Recommendation

EMI lecturers should remember that the accumulated knowledge of written receptive vocabulary—the explicit knowledge of L1 meaning and L2 form connections—alone may not indicate students’ future readiness for their success in EMI. Alongside traditional vocabulary measures highlighting the explicit knowledge, gauging students’ implicit knowledge or the ability to use L2 words fluently through contextualized tasks may provide additional information indicating whether they will be successful in performing EMI tasks. Regarding vocabulary teaching, EMI preparatory or EAP courses should focus on developing fluency (i.e., implicit knowledge of vocabulary) rather than simply increasing decontextualized vocabulary knowledge through direct teaching. Learners could benefit from activities requiring oral production of known words under increasing time pressure such as the 4/3/2 activity (see Webb & Nation, 2017, pp. 88-89). EMI teachers should also note that unsatisfactory experience in L2 production activities may result in students’ lack of confidence in their vocabulary knowledge. Focusing on productive vocabulary bydevoting class time to meaning-focused output and fluency-development activities (Nation, 2013) may be effective for students to gain a positive experience in performing production tasks. Another point worth noting is that teachers cannot estimate students’ aural vocabulary size from their written vocabulary size. In this study, of all students (N = 35) who mastered knowledge of the most frequent 2,000 word families in written form, more than half of them failed to master the aural form of the same frequency level. The mismatch may be common among EFL settings where learners have not been well trained in oral skills. This could be a serious obstacle to academic success, given that 4,000-5,000 word families are necessary to comprehend academic spoken English (Dang & Webb, 2014). A possible solution to this issue is to prepare students to get used to aural forms of vocabulary in an EAP course. It is effective to have students listen to short video clips of a content lecture with focus on a variety of phonetic forms in different linguistic or informational contexts (e.g., words in a prominent or non-prominent position, words appearing as new or given information) or produced by different lecturers with various linguistic backgrounds (Brinton & Jensen, 2002). Thus, students’ understanding of aural forms of content words will be facilitated, which helps develop a robust phonological form of content vocabulary (Bent, Kewley-Port, & Ferguson, 2010). EMI teachers should also keep in mind that students with inadequate aural vocabulary will appreciate scaffolding from visual aids, code-switching, and constant interaction to ensure their understanding of the lecture. Finally, lack of technical vocabulary poses challenges to EMI students. Selecting course textbooks with a glossary of technical terms may be helpful for students. It may also be effective to teach relevant specialized words using existing lists of
technical vocabulary in EAP classes before content courses start (see Webb & Nation, 2017, p. 16 for references to existing technical-word lists for different academic fields).
References


Harada, T. (2017). Developing a content-based English as a foreign language program: Needs analysis and curriculum design at the university level. In M. A. Snow & D. M. Brinton (Eds.), The content-based classroom: New perspectives on...
integrating language and content (pp. 37-52). Ann Arbor, MI: University of Michigan Press.


Footnotes

1 Eleven out of 300 target items (3.7%) in VLT and LVLT overlapped: abandon, accumulate, circus, diminish, independent, jump, professional, pursue, retain, rigid, and solution.

2 “CLIL” here refers to the title of the content course which 16 out of the 35 participants took at the time of data collection. The remaining students attended a range of EMI courses (e.g., Advanced Phonetics, Undergraduate Seminar on Bilingual and Immersion Education).

3 Negative correlations were also confirmed between each sub-section of the VLT scores and each subcategory item of the four-skill perception scores (see Appendix 8).