



Deposited via The University of Leeds.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/237017/>

Version: Accepted Version

Article:

Lu, C. and Dang, T.N.Y. (2023) Effect of L2 exposure, length of study, and L2 proficiency on EFL learners' receptive knowledge of form–meaning connection and collocations of high-frequency words. *Language Teaching Research*. ISSN: 1362-1688

<https://doi.org/10.1177/13621688231155820>

© The Author(s) 2023. This is an author produced version of an article published in *Language Teaching Research*. Uploaded in accordance with the publisher's self-archiving policy.

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

Effect of L2 Exposure, Length of Study, and L2 Proficiency on EFL Learners' Receptive Knowledge of Form-Meaning Connection and Collocations of High-Frequency Words

Cailing Lu

Shanghai Jiao Tong University, China

Thi Ngoc Yen Dang

University of Leeds, UK

Abstract

High-frequency words are single-words at the first three 1,000-word frequency levels. No studies have explored the impact of current second language (L2) exposure, length of study, and L2 proficiency on the receptive knowledge of high-frequency words of EFL learners from non-cognate L1 background. To address this gap, this study used the Updated Vocabulary Levels Test (Webb et al., 2017) and the Receptive Collocation Test (Nguyen & Webb, 2017) to measure the receptive knowledge of form-meaning connection and collocations of high-frequency words of 201 Chinese EFL postgraduate students. It also surveyed these students' amount of current L2 exposure, length of study, and L2 proficiency. The results showed that, on average, the participants had 20.22 hours of exposure to various types of L2 input per week in and out of class. A series of single linear mixed-effects model analyses revealed that their current L2 exposure and length of study did not seem to have an impact on their knowledge of form-meaning connection of high-frequency words nor their collocational knowledge of these words overall. Yet, L2 proficiency significantly affected their knowledge of form-meaning connection and collocations of the 1st 1,000 words. Knowledge of form-meaning connection also contributed to the collocational knowledge. Pedagogical implications for vocabulary researchers and practitioners were provided.

Keywords

high-frequency vocabulary; collocations; Chinese postgraduates; language exposure; L2 proficiency

1. Introduction

High-frequency words are single-words at the first three 1,000-word frequency levels, or the most frequent 3,000 words (e.g., *money*, *advantage*, *conflict*) (Schmitt & Schmitt, 2014). A good knowledge of high-frequency words is crucial for successful use of English as a foreign language (EFL) (Nation, 2022). Such knowledge involves both knowledge of form-meaning connection of high-frequency words and knowledge of collocations¹ of these words (e.g., *make money*, *take advantage*, *avoid conflict*) (Nation, 2022). In vocabulary research, the former dimension is often referred to as vocabulary breadth while the latter belongs to vocabulary depth (Nation, 2022; Read, 2000). Despite the importance of having a broad and deep knowledge of high-frequency words, earlier research showed that a considerable number of learners in various EFL contexts had insufficient knowledge of these lexical items (e.g., Dang, 2020a, 2020b; Henriksen & Danelund, 2015; Lu & Dang, 2022; Matthews & Cheng, 2015; Nguyen & Webb, 2017; Olmos, 2009; Sun & Dang, 2020). To better support EFL learners' development of high-frequency word knowledge, it is important to explore factors that may affect their breadth and depth of knowledge of these words.

Despite this need, to the best of our knowledge, only Nguyen and Webb (2017) have explicitly examined the relationship between EFL learners' knowledge of high-frequency words and the factors affecting their knowledge of these lexical items. However, they only examined factors influencing collocational knowledge of high-frequency words. Investigating factors affecting knowledge of form-meaning connection of high-frequency words would also be useful, because it is the basic foundation for the acquisition of other aspects of vocabulary knowledge (Webb & Nation, 2017). Such investigation would provide researchers with a better understanding of various factors that influence the breadth and depth of high-frequency word knowledge. Moreover, Nguyen and Webb focused on word-related factors (node word frequency, collocation frequency, association strength, congruency, part of speech). Yet learner-related factors such as L2 exposure, length of study, and L2 proficiency may play a role in vocabulary learning (Peters, 2020).

Meanwhile, studies examining the effects of L2 exposure, length of study, and L2 proficiency on EFL learners' vocabulary knowledge did not specifically investigate high-frequency words. They focused on either learners' knowledge of the most frequent 5,000 words, 6,000 words, or 10,000

words (e.g., González Fernández, 2021; González Fernández & Schmitt, 2015; Peters, 2018; Puimège & Peters, 2019). Additionally, most of the examined EFL learners were from cognate L1 backgrounds²: Spanish L1 speakers in Spain (González Fernández & Schmitt, 2015) and Dutch L1 speakers in Belgium (Peters, 2018; Peters et al., 2019; Puimège & Peters, 2019). To date, only González Fernández (2021) has involved both EFL learners from cognate L1 backgrounds (Spanish) and non-cognate L1 backgrounds (Chinese), but she considered these learners as a whole group and did not further examine the impact of L2 exposure on vocabulary knowledge of sub-groups of learners. Research with EFL learners from non-cognate L1 backgrounds would be beneficial. Unlike those from cognate L1 backgrounds, learners from non-cognate L1 backgrounds do not have the advantages brought by the similarities between English and their L1s when learning English (De Wilde et al., 2019; Peters, 2020).

The present study was conducted to fill in these gaps. Focusing on EFL postgraduate students from a non-cognate L1 background (Chinese), this study aims to (a) corroborate Nguyen and Webb's (2017) findings for the relationship between these learners' receptive knowledge of form-meaning connection of high-frequency words and their collocations, and (b) extend their findings by exploring the effect of L2 proficiency, amount of current L2 exposure, and length of study English on the learners' receptive knowledge of form-meaning connection and collocations of these lexical items. This study would help to find factors that may affect EFL learners' knowledge of high-frequency words. This would, in turn, provide useful implications for vocabulary researchers and practitioners to create ideal learning conditions to optimize EFL learners' development of high-frequency words in terms of breadth and depth.

2. Background

2.1. EFL learners' knowledge of high-frequency words

A good knowledge of high-frequency words is essential for EFL learners (Nation, 2022). First, high-frequency words make up a substantial proportion of the vocabulary (75%-91%) in written and spoken English texts (Dang & Webb, 2020). Knowledge of these words is fundamental for comprehension of any discourse types. Second, L2 learners tend to acquire vocabulary according to frequency, a good knowledge of high-frequency words would provide a firm foundation for the acquisition of words at lower frequency levels (Ellis, 2002). A reasonable number of studies

have measured EFL learners' knowledge of high-frequency words. They consistently reported that learners from various EFL contexts had insufficient knowledge of high-frequency words after a long period of formal English instruction (e.g., Dang, 2020a, 2020b; Henriksen & Danelund, 2015; Lu & Dang, 2022; Matthews & Cheng, 2015; Olmos, 2009; Sun & Dang, 2020). However, these studies only measured the vocabulary breadth (i.e., knowledge of form-meaning connection). The depth of knowledge is also important (Read, 2000). Such knowledge involves many aspects (Nation, 2022). One important aspect is collocational knowledge. It is the basis of fluent language use (Siyanova-Chanturia & Pecillia-Sanchez, 2018). Moreover, research has shown that EFL learners' collocational knowledge tends to lag behind other aspects of vocabulary knowledge (e.g., synonyms, word parts) (Read & Dang, 2022). Yet research measuring EFL learners' collocational knowledge mainly examined knowledge of collocations of words from various frequency levels (e.g., González-Fernández, 2021; González Fernández & Schmitt, 2015) rather than focusing specifically on collocations of high-frequency words. In fact, research examining both the breadth and depth of high-frequency word knowledge is limited in number.

To the best of our knowledge, only Nguyen and Webb (2017) have focused on EFL learners' receptive knowledge of form-meaning connection and collocations of high-frequency words, as well as their relationship. They found that after at least seven years of studying English, university English majors in Vietnam knew 93.57%, 77.83%, and 65.03% of the high-frequency words from the 1st, 2nd, and 3rd 1,000-word levels receptively in terms of form-meaning connection. Meanwhile, they knew only 57.86%, 43.35%, and 35.28% of the collocations of high-frequency words from the 1st, 2nd, and 3rd 1,000-word levels, respectively. This indicated that the participants' development of knowledge of form, meaning, and collocations of high-frequency words was slow. In addition, Nguyen and Webb found a significant relationship between EFL learners' receptive knowledge of form-meaning connection of high-frequency words and their collocations. Further, node word frequency, strength of association, and congruency of the components of collocations appeared to make a significant contribution to the participants' collocational knowledge, but part of speech did not. While Nguyen and Webb (2017) provided useful insights into EFL learners' receptive knowledge of high-frequency words, it was limited to one EFL population (i.e., undergraduate students in Vietnam). Thus, it is unclear whether the same pattern would be found with EFL learners at different educational

levels or from different L1 backgrounds. Moreover, Nguyen and Webb only examined the effect of word-related factors on vocabulary knowledge. Learner-related factors such as L2 exposure, length of study, and L2 proficiency may also have a certain impact (Peters, 2020). Also, they only focused on the effect on collocational knowledge of high-frequency words. Examining the impact on knowledge of both form-meaning connection and collocations would provide a more comprehensive picture of various factors that influence the breadth and depth of high-frequency word knowledge. Given this, the next three sections will review research which investigated learner-related factors that may affect vocabulary learning in turn.

2.2. The effect of EFL learners' exposure to L2 input on their vocabulary knowledge

The positive effect of EFL learners' L2 exposure (e.g., reading, watching TV, playing computer games) on vocabulary knowledge has been demonstrated by several empirical studies. Most of these studies focused on knowledge of form and meaning. Muñoz (2011) examined the impact of EFL learners' L2 exposure (examined by a self-reported questionnaire) on 162 Spanish EFL students' vocabulary knowledge at a university. L2 exposure in this study included all the exposure to the L2 since the learners started learning English. That means it included both current L2 exposure and length of study, and both in-class and out-of-class exposure. The results showed that length of study and recent in-class and out-of-class L2 exposure positively correlated with their receptive knowledge of the most frequent 5,000 and 10,000 English words at the level of form-meaning connection. This positive relationship between L2 exposure and vocabulary knowledge was later corroborated by Peters and colleagues, who conducted a series of studies investigating the effects of out-of-class language exposure on vocabulary knowledge with EFL learners at different educational levels in the Dutch-speaking region of Belgium. Peters' (2018) study with secondary school students and undergraduate students revealed that listening to songs, watching movies/TV programs with and without subtitles, playing games and browsing the Internet were the most common activities whereas reading books and magazines was the least frequent. She also found that these learners' out-of-class L2 exposure (measured by a self-reported questionnaire) significantly correlated with their receptive knowledge of the form-meaning connection of the most frequent 5,000 words. Among the examined kinds of input, browsing the Internet had the strongest relationship with learners' vocabulary knowledge. Next came watching movies/TV programs without subtitles, and reading magazines, and then reading books. There was no significant relationship between watching movies/TV programs

with subtitles and playing computer games, and vocabulary knowledge while a negative relationship was found in the case of listening to songs.

Peters et al.'s (2019) subsequent study investigated the impact of out-of-school exposure (measured by a self-reported questionnaire) to foreign language input on secondary school students and undergraduate students' receptive vocabulary knowledge in two foreign languages – French and English. The results showed that the participants' out-of-class exposure to French, which is their first foreign language, did not significantly correlated with their receptive knowledge of the form-meaning connection of the most frequent 5,000 words in French. In contrast, these participants' out-of-class exposure to English positively contributed to their knowledge of the most frequent 5,000 words in English. Similar to the previous study, this study revealed that listening to songs, watching movies/TV programs with and without subtitles, playing computer games, and visiting websites in English were the participants' main sources of out-of-class exposure while reading written media was not. Among these sources, only online activities significantly affected the participants' vocabulary knowledge while the use of English with friends/families, and other audiovisual materials with and without subtitles, and written media did not. A fairly similar result pattern was also found with primary school students from Belgium (Puimège & Peters, 2019).

Compared to research on knowledge of form-meaning connection, fewer studies have investigated the effect of L2 input on EFL learners' collocational knowledge. González-Fernández and Schmitt's (2015) study with adult Spanish EFL learners found a significant relationship between these participants' productive knowledge of collocations of the most frequent 5,000 words and their exposure to English through different kinds of out-of-class input (measured by a self-reported questionnaire). Among the individual sources, reading, watching films/videos/TV programs, social networks, and immersion in English speaking countries significantly affected the vocabulary knowledge, but listening to music did not. Similarly, González-Fernández's (2021) study with Spanish and Chinese EFL learners showed a facilitative effect of L2 exposure on the participants' receptive knowledge of form-meaning connection and collocations sampled from the most frequent 10,000 words. Specifically, reading, viewing TV, and using social network were reported to have significant effects on the participants' breadth and depth of vocabulary knowledge. Interestingly, this study also found a significant but small

correlation between listening to music and learners' receptive knowledge of form-meaning connection.

Previous studies have provided important insights into the effect of EFL learners' L2 exposure on their vocabulary knowledge. However, several areas need further attention. First, most earlier studies examined learners from cognate L1 backgrounds. The only exception is González Fernández's (2021), but that study combined both groups of learners in the analysis rather than distinguish between learners from cognate and non-cognate L1 backgrounds. Consequently, the effect of the amount of L2 exposure on the vocabulary knowledge of non-cognate L1 students solely still remains unknown. Additionally, except for Muñoz (2011), previous research mainly focused on out-of-class exposure while in-class exposure is also an important source of input in many EFL contexts (e.g., Alsaif & Milton, 2012; Jordan & Gray, 2019). Third, previous studies investigated knowledge of form, meaning, and collocations of the most frequent 5,000 words, 6,000 words, and 10,000 words. No studies have focused specifically on those aspects of knowledge of high-frequency words. This is surprising given the significant role of high-frequency words for EFL learners (Nation, 2022).

2.3. The effect of length of study on EFL learners' vocabulary knowledge

Another factor that has attracted considerable research interests on EFL learners' vocabulary knowledge is the length of study. To the best of our knowledge, four studies have investigated this factor, three of which focused on knowledge of form and meaning. Muñoz's (2011) study, as reviewed in Section 2.2, reported that length of study positively correlated with their receptive vocabulary knowledge. The positive effect of length of study on EFL learners' vocabulary knowledge was then extended to Flemish EFL learners at secondary school and university levels in Belgium by Peters (2018) and Peters et al.'s (2019), who also used self-reported questionnaires to measure the length of study. However, these three studies only examined the effect of length of study on EFL learners' knowledge of form-meaning connection. Moreover, they only examined secondary school students and undergraduate students. Postgraduate students are likely to have a longer time of studying English and perhaps are likely to be exposed to a larger amount of L2 input. Therefore, the effect of length of study on their vocabulary knowledge might be different from the results of the other EFL population.

Only González Fernández and Schmitt (2015) have investigated the effects of the length of study on collocational knowledge. Similar to previous research, their adult Spanish EFL learners were asked to complete a self-reported questionnaire and a lexical test. This study revealed that length of study was also an important predictor of learners' knowledge of collocations of the most frequent 5,000 words. However, González Fernández and Schmitt (2015) centred on learners' productive collocational knowledge. The extent to which length of study influences EFL learners' receptive collocational knowledge remains to be investigated. Taken together, the above studies either investigated the effect of length of study on cognate L1 EFL learners' knowledge of form-meaning connection or collocations. No single studies have examined the effect of length of study on non-cognate L1 EFL learners' receptive knowledge of high-frequency words of various dimensions – form, meaning, and collocations, which indicates a clear need for future research.

2.4. The effect of L2 proficiency on EFL learners' vocabulary knowledge

Research into the effect of L2 proficiency on vocabulary knowledge has also received considerable attention. Gui (2015) has investigated the correlation between 96 Chinese EFL learners' vocabulary breadth (i.e., knowledge of form-meaning connection) and L2 proficiency, which was represented by their scores on the College English Test Band 4 and 6. The College English Test is a national English test for non-English major university students in China. Band 4 is for students at a relatively lower proficiency level whereas Band 6 is for those with a higher proficiency level who have passed Band 4. This study revealed that Chinese EFL learners' vocabulary breadth significantly correlated with their scores on the College English Test Band 4, but not College English Test Band 6. However, Gui (2015) did not investigate the relationship between the participants' collocational knowledge and L2 proficiency.

To the best of our knowledge, only González-Fernández and Schmitt (2015) has looked at this aspect from the perspective of collocational knowledge. Examining Spanish EFL adult learners who self-rated their L2 proficiency as beginner, intermediate, or advanced, they found that the participants' L2 proficiency significantly contributed to their knowledge of collocations of the most frequent 5,000 words. However, as González Fernández and Schmitt (2015) acknowledged, their findings should be interpreted with caution. The L2 proficiency was self-rated by the participants themselves in a multiple-choice questionnaire; thus, the rating can be subjective.

Moreover, González-Fernández and Schmitt did not specify the educational levels of their participants. Consequently, it is unclear how L2 proficiency may affect the collocational knowledge of EFL learners on a specific educational level (e.g., primary school, postgraduate study). To date, no studies have investigated the relationship between non-cognate EFL learners' L2 proficiency and their knowledge of form, meaning and collocations, which indicates a gap in vocabulary research. Such research should use more objective measures to rate the participants' L2 proficiency.

2.5. Research questions

To address the gaps mentioned in the literature review, this study investigated EFL postgraduate students from a non-cognate L1 background (Chinese) and aimed to address the following research questions:

1. To what extent is Chinese EFL postgraduates' receptive knowledge of form-meaning connection of high-frequency words affected by the amount of current exposure to English, length of study, and L2 proficiency?
2. To what extent is these students' receptive collocational knowledge of high-frequency words affected by their receptive knowledge of form-meaning connection of high-frequency words, the amount of current exposure to English, length of study, and L2 proficiency?

3. Methodology

3.1. Participants

Originally, 238 first-year postgraduate students at a university in the east of China participated in the study. These students voluntarily participated in the present study following the authors' institution ethical practices. Only the data of 201 students (141 males and 60 females, aged between 21 to 27 years old) who completed all steps in Section 3.3 were used for the analysis. All participants shared features of postgraduates in China and many other EFL contexts. They had learned English as a foreign language in their home country for at least 10 years from primary school, middle school, high school, to undergraduate study. On average, those learners were at a proficiency level of just passing the College English Test Band 4, which is equivalent to the IELTS score of 5. To be admitted to the postgraduate program, the participants had to take

the National Postgraduate Entrance Examination in China. For the English subject of this examination, they needed to score at least 37 out of 100. By the time this research was conducted, the participants were taking a compulsory English for Academic Purposes course at the beginning of their three-year Master's study. This was an intensive course delivered by Chinese L1 lecturers with high English language proficiency level (relevant to IELTS 7 or higher). It ran for three hours each week and lasted for 16 weeks. Apart from in-class contact time, the students were expected to spend about three hours per week on out-of-class activities (e.g., reading academic texts, preparing oral presentations).

3.2. Instruments

Updated Vocabulary Levels Test and Receptive Collocation Test

The 1st, 2nd and 3rd 1,000-word levels of the Updated Vocabulary Levels Test (UFLT) (Webb et al., 2017) were administered to measure the participants' receptive knowledge of high-frequency words at the level of form-meaning recognition. The test is in the matching format in which students have to match the words with their relevant definitions. Each level has 30 items. According to Webb et al. (2017), to master each level, students need to get 29 out of 30 correct answers per level (see Figure 1 for the example of the test item).

	cable	section	sheet	site	staff	tank
part						
place						
something to cover a bed						

FIGURE 1. Examples of the Update Vocabulary Levels Test items

The Receptive Collocation Test (RCT) (Nguyen & Webb, 2017) was administrated to investigate participants' collocational knowledge of high-frequency words. This test is in a multiple-choice format (see Figure 2 for the example of the test item). It has three levels, each of which measures knowledge of verb–noun and adjective–noun collocations of the 1st, 2nd, and 3rd 1,000 words, respectively. As shown in Figure 2, the prompt word in each collocation was a noun with a verb or an adjective to be chosen as its collocate.

13. tax	<i>a. charge</i>	<i>b. pay</i>	<i>c. run</i>	<i>d. share</i>
14. court	<i>a. fill</i>	<i>b. hold</i>	<i>c. take</i>	<i>d. turn</i>
15. comfort	<i>a. make</i>	<i>b. lose</i>	<i>c. put</i>	<i>d. take</i>

FIGURE 2. Examples of the RCT items

The UVLT was chosen for three reasons. First, it was based on Nation's (2012) BNC/COCA word lists, which contain the most useful high-frequency words for EFL learners from the perspective of corpora, learners, and teachers (Dang et al., 2020). Second, the UVLT has been carefully validated by Webb et al. (2017) and widely employed to measure EFL learners' receptive knowledge of form-meaning connection (e.g., Nguyen, 2020; Sun & Dang, 2020). Third, the UVLT can provide a more accurate assessment of L2 learners' receptive knowledge of general high-frequency vocabulary at each 1,000-word level, while previous VLT versions (Nation, 1983; Schmitt et al., 2001) were less sensitive because they clumped several 1,000-word frequency levels together.

The RCT was chosen because it was developed in a similar way as the UVLT which involved sampling test items based on Nation's (2012) corpus-based BNC/COCA lists that were divided into 1,000-word frequency bands. Both tests were developed with the assumption that performance on a sample of words from a particular frequency level reflects knowledge of all the words in that band. With both tests derived from the same BNC/COCA frequency lists, the measures are well suited to the exploration of the relationship between the learners' knowledge of single words and collocations. Moreover, this test has been used by Nguyen and Webb to measure the collocational knowledge of Vietnamese EFL university students who had a fairly similar background as the participants in the present study.

The UVLT and RCT have gone through a certain rigorous process of development and have been used by previous research to measure EFL learners' vocabulary knowledge. To some extent, they are reliable instruments to measure EFL learners' knowledge of form-meaning connection and collocations of high-frequency words. However, ten words tested in the UVLT (*note, war, work, feature, knowledge, population, intelligence, poverty, wealth, conflict*) are also components of 10 collocations tested in the RCT (*take **note**, cold **war**, hard **work**, important **feature**, gain **knowledge**, general **population**, gather **intelligence**, reduce **poverty**, create **wealth**, avoid **conflict***). As the RCT was administered after the UVLT, there is a concern that completion of the UVLT would lead to a significant improvement on the participants' RCT scores. However, the chance for this improvement to happen is fairly small. First, given the large number of items in the two tests, it is unlikely that the participants could notice and learn these overlapping items

from taking the test. Second, each overlapping word is only one component of the target collocation. Collocational knowledge tends to lag behind knowledge of form-meaning connections (e.g., Akbarian, 2010; Nguyen & Webb, 2017); therefore, knowing the form-meaning connection of these UVLT words does not necessarily guarantee that the participants could successfully select the words they collocate with in the RCT.

Language Learning Experience Questionnaire

A language learning experience questionnaire was used to survey the participants' L2 proficiency, amount of current exposure to English, and length of studying English (see Appendix 1 for the full survey). The L2 proficiency was represented by the participants' score on the English test of the National Postgraduate Entrance Examination. According to the official guide³ published by the Chinese Ministry of Education Examination Centre (2018), this test consists of three parts: 10% of English knowledge (i.e., vocabulary, syntax, and grammar) application, 60% of reading comprehension, and 30% of essay writing. Comparable to the English test of the College Entrance Examination for undergraduate students, the English test of the National Postgraduate Entrance Examination is a national-wide and high-stake test. It is held annually as part of the National Postgraduate Entrance Examination. Scores on this test has long been used to measure students' English proficiency as part of the admission requirements for almost all postgraduate programs in mainland China. As a result, this test was considered as being reliable in representing the postgraduates' L2 proficiency. Moreover, care was taken in the data collection to ensure the participants understand the research purpose and complete the questionnaire seriously. Additionally, results of the Shapiro-Wilk test of normality ($p = 0.09$) indicated that the test scores reported by the participants normally distributed. This suggests that it is unlikely that the participants were dishonest when reporting their test scores and confirms the reliability of the test scores to some extent.

The amount of exposure to English was measured in terms of the cumulative number of hours per week that the learners currently got exposure from different kinds of input. We focused on students' current exposure to English because the participants may not remember details about their previous exposure to English well. Examining their current exposure also allowed us to have a good understanding of the students' current experiences and offer relevant pedagogical implications for teachers and students in the examined EAP programs. The length of study was

counted since the participants started learning English (from primary school in most cases); extracurricular English courses were also counted. If the students had a break between their undergraduate and postgraduate studies, and they did not take any curricular and extracurricular English courses, that period of time was not counted.

We chose the self-reported questionnaire format to collect information about current L2 exposure, L2 proficiency, and length of study because this format has been widely used in previous studies to examine the effects of these factors on vocabulary knowledge (e.g., González-Fernández, 2021; González-Fernández & Schmitt, 2015; Muñoz, 2011; Peters, 2018; Peters et al, 2019). Following the same method would enable us to compare the results of our study with those from previous studies. Using self-reported questionnaires also has practical values. Questionnaires allow us to collect a large sample in a short time and they are easy and cheap to administer (Cohen et al., 2018). However, learners may have different interpretations of the wording in the questions or they may not answer the questions honestly in self-reported questionnaires (Dörnyei, 2007). Therefore, when designing and administering the questionnaire, we have taken several steps to minimize these limitations as much as possible. In terms of designing the questionnaire, we adapted it from Peters' (2018) questionnaire rather than creating it from scratch. Peters' questionnaire was based on the European Survey of Language Competences and had been extensively piloted and used in previous studies by Peters and colleagues (e.g., Peters, 2018; Peters et al., 2019). However, Peters' questionnaire only examined learners' exposure to English outside classroom. Classroom instruction is also important in vocabulary learning in EFL context (e.g., Dang & Webb, 2020; Jin & Webb, 2020). Therefore, we also added items to explore learners' exposure to English inside English language classes. After both authors were happy with the first draft of the questionnaire, we asked eight students from similar backgrounds as the target participants to complete the questionnaire. Based on their feedback, we revised the problematic items so that all questions were clear and easy to understand. Then, we piloted the revised questionnaire with 50 students from the same course as the target participants. The pilot data showed that the questionnaire had good internal consistency (Cronbach alpha coefficient = .79). Analysis of the main study data also indicated that the questionnaire had good internal consistency (Cronbach alpha coefficient = .76).

3.3. Procedure

This study was conducted at the beginning of the participants' postgraduate study in their English classes. First, the participants completed the UVLT, then the RCT, and finally the language learning experience questionnaire. The tests and questionnaire were in a pencil-and-paper format. It took the participants about 25 minutes, 30 minutes, and 10 minutes to complete the UVLT, the RCT, and the questionnaire, respectively. To avoid fatigue effects, there were breaks between the tests.

Care was also taken in the data collection to ensure the participants had a good understanding of the purpose and content of the questionnaire. The first author administered the questionnaire to students in their class time and supervised the process to ensure that the participants answered the questionnaires seriously. Throughout this process, she informed the participants that the purpose of the questionnaire was for research only and their answers would not affect their final scores on the EAP course. She also explained in detail the definitions of the amount of current exposure to English and length of study adopted in our study and addressed any queries from the participants. All instructions were given to the participants in their L1 to avoid misunderstanding.

3.4. Data analysis

The UVLT and RCT were scored dichotomously. A correct response was given 1 point whereas an incorrect response was given 0. With respect to the questionnaire, following previous studies (González Fernández & Schmitt, 2015; Puimège & Peters, 2019), items related to the amount of exposure to English in the questionnaire was coded following the protocol in Table 1.

TABLE 1

Questionnaire coding

Term in questionnaire	Code
Never	0
≤1 hour	1
>1 to 2 hours	2
>2 to 3 hours	3

>3 to 4 hours	4
>4 to 5 hours	5
>5 to 6 hours	6
>6 to 7 hours	7
>7 to 8 hours	8
>8 to 9 hours	9
>9 to 10 hours	10
> 10 hours	11

Note: The unit of exposure is hours per week.

Research question 1 asked about the effect of current exposure to English, length of study, and L2 proficiency on receptive knowledge of form-meaning connection of high-frequency words. To answer this question, we conducted a series of single linear mixed-effects models with the lme4 package in R. In these models, the three fixed factors were exposure to English, length of study, and L2 proficiency while the random factor was the participants' class. The UVLT score was the dependent variable. We also conducted similar series of single linear mixed-effects models to answer Research question 2, which examined the effect of the amount of current exposure to English, length of study, L2 proficiency, and knowledge of form-meaning connection of high-frequency words on receptive knowledge of collocations of high-frequency words. However, in these models, the RCT score was the dependent variable, and the UVLT score was added as the fourth fixed factor. The VIF scores in all cases were between 1.0 and 1.9, indicating no issues of multicollinearities. Mixed-effects models were chosen for the analysis because of their increasingly popularity in second/foreign language research. This statistical technique allows researchers to include a wide range of independent variables of their interests in a single model as fixed factors; at the same time, it takes into consideration any potential variances due to individual differences through the inclusion of random effects (Cummings, 2012).

4. Results

4.1. Receptive knowledge of form-meaning connection and collocations of high-frequency words

Table 2 presents the descriptive statistics of the participants' scores on the UVLT and the RCT. In both tests, their scores on the 1st 1,000-word level were the highest. Next came their scores on the 2nd 1,000-word level. Their scores on the 3rd 1,000-word level were the lowest. One-way repeated measure ANOVA showed that these differences were significant at $p < 0.05$. When the 29-out-of-30 criterion for mastery of a 1,000-word frequency level (Webb et al., 2017) was applied, the results showed that the participants had nearly mastered the 1st 1,000 words, but not the 2nd and 3rd 1,000 words. The participants' collocational knowledge of the high-frequency words was also limited. The RCT scores indicated that on average the participants knew only nearly 58% of the 1st 1,000-word level collocations, nearly 46% of the 2nd 1,000-word level collocations, and nearly 35% of the 3rd 1,000-word level collocations.

TABLE 2

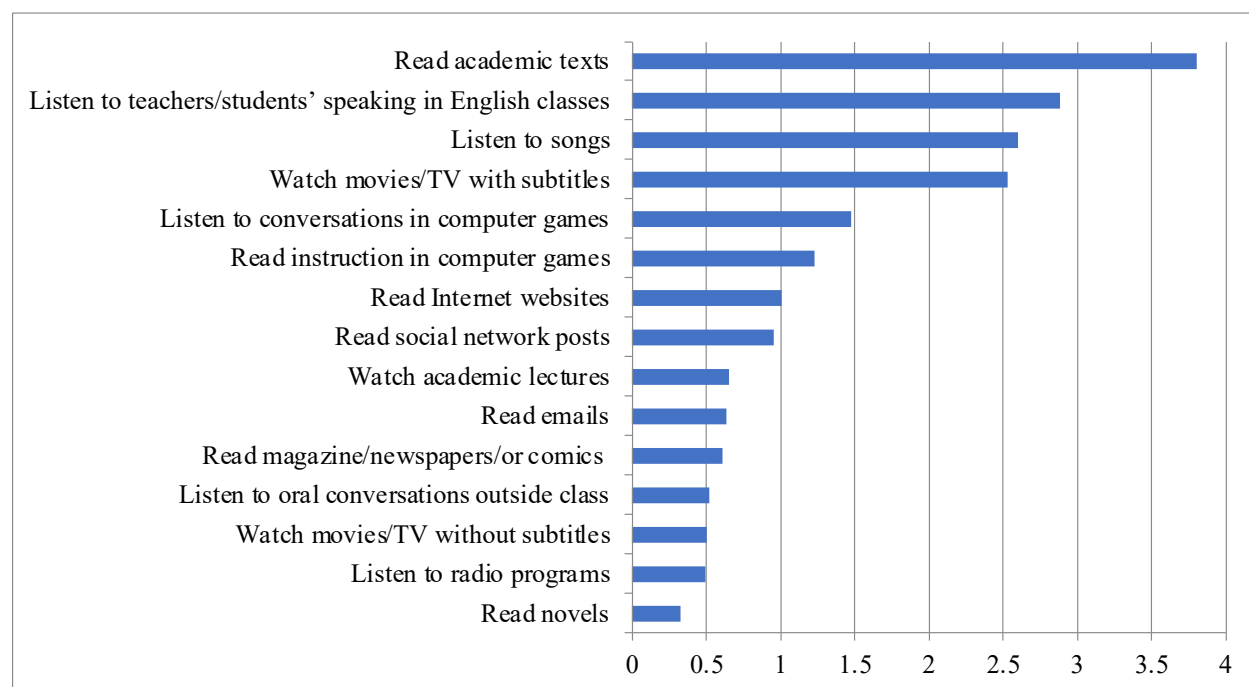
Mean (SD) of the UVLT and RCT scores

	Number of correct responses		Percentage of correct responses (%)	
	UVLT	RCT	UVLT	RCT
Frequency levels	(single words)	(collocations)	(single words)	(collocations)
1 st 1,000-word level	28.94 (1.59)	34.72 (4.4)	96.47 (5.27)	57.87 (7.34)
2 nd 1,000-word level	24.84 (3.36)	27.38 (5.44)	82.80 (11.18)	45.64 (9.06)
3 rd 1,000-word level	21.49 (4.41)	20.87 (6.60)	71.64 (14.69)	34.78 (11.00)
All levels combined	75.27 (7.78)	82.98 (12.64)	83.64 (8.67)	46.10 (7.02)

4.2. Amount of current L2 exposure, length of study, and L2 proficiency

Figure 3 presents the descriptive statistics of the average amount of time that the participants were exposed to each kind of input per week. For all types of input, the code is below 4 out of 11. It means that regardless of the types of input, the participants spent no more than 4 hours per week in each kind of input. Reading academic texts is the most frequent activity. Listening to teachers/students' speaking in English language class ranks the second, followed by listening to songs and watching movies/TV programs with subtitles. Next came listening to conversations in computer games, reading instructions in computer games, reading Internet websites and social

network posts. After that are other activities (e.g., watching academic lectures, reading emails). The participants spent the least time on reading novels. On average, the participants had 20.22 hours exposure to input per week ($SD = 10.75$) in and out of class. The analysis of the questionnaire also showed that, on average, the participants studied English for 12.62 years ($Max = 20$, $Min = 10$, $SD = 2.10$) and they scored 57.59 out of 100 ($Max = 83$, $Min = 37$, $SD = 9.49$) in the English test of the National Postgraduate Entrance Examination.



Note: The unit of exposure is hours per week.

FIGURE 3. Participants' amount of time exposed to English input per week

4.3. The relationship between receptive knowledge of form-meaning connection of high-frequency words and three factors

In answer to Research question 1, Table 3 shows that L2 proficiency had a positive significant contribution to the participants' receptive knowledge of the form-meaning connection of high-frequency words at $p < 0.001$. It means that L2 proficiency appeared to be a good indicator of the participants' knowledge of form-meaning connection of high-frequency words as a whole. In contrast, the amount of composite exposure to L2 and the length of study did not make any significant contribution to the participants' receptive knowledge of form-meaning connection.

Similar results were found in the case of knowledge of high-frequency words at the 1st, 2nd, and 3rd 1,000-word levels, respectively (see Appendices 2-4). L2 proficiency significantly and positively affected the participants' receptive knowledge of the form-meaning connection of high-frequency words as a whole as well as words at each 1,000-word level, but the composite exposure to L2 and the length of study did not.

TABLE 3

Results of the single linear mixed-effects model analysis for knowledge of form-meaning connection of high-frequency words (1st, 2nd, 3rd 1,000-word levels combined)

	Estimate	Std. Error	df	t value	Pr(> t)
(Intercept)	4.12	0.07	136.7	63.07	< 0.0001***
Composite exposure to L2	0.004	0.01	196	0.40	0.69
Proficiency	0.003	0.0008	119.7	3.58	0.0005***
Length of study	0.002	0.004	183.2	0.43	0.66

*** sig. at $p < 0.001$, ** sig. at $p < 0.01$, * sig. at $p < 0.05$. Note: Composite score refers to the sum of all types of exposure.

Table 4 shows that when the effects of each source of L2 input, L2 proficiency, and length of study on the participants' receptive knowledge of form-meaning connection of high-frequency words (1st, 2nd, and 3rd 1000-word levels combined) was examined, L2 proficiency had a positive, significant effect, but the length of study did not. Among the different sources of L2 input, only reading people's posts in English on social networks significantly affected the participants' receptive knowledge of form-meaning connection of high-frequency words ($p = 0.03$). Appendices 5-7 present the results about the effects of each source of L2 input, L2 proficiency, and length of study on the participants' knowledge of the form-meaning connection of words at the 1st, 2nd, and 3rd 1,000-word levels, respectively. L2 proficiency significantly affected the learners' vocabulary knowledge at each frequency level, but length of study did not. Meanwhile, none of the sources of L2 input affected the participants' knowledge of the 1st and 2nd 1,000 words. However, reading Internet websites and reading people's posts on social networks both contributed to the participants' knowledge of the 3rd 1,000 words ($p = 0.04$, see Appendix 7 for details).

TABLE 4

Results of the single linear mixed-effects model analysis for knowledge of form-meaning connection of high-frequency words (1st, 2nd, 3rd 1,000-word levels combined) (individual sources of L2 input)

	Estimate	Std. Error	df	t value	Pr(> t)
(Intercept)	4.1	0.07	140	60.54	< 0.0001**
Proficiency	0.003	0.001	100	3.72	0.0003***
Length of study	0.002	0.004	150.9	0.51	0.61
Listening to student/teacher talks in English classes	-0.0003	0.007	158.9	-0.06	0.96
Listening to songs	-0.004	0.004	180.7	-1.16	0.25
Listening to radio	-0.004	0.01	180.9	-0.47	0.64
Listening to general conversations in English outside class in person or virtually	0.002	0.01	180.7	0.21	0.83
Listening to instruction and/or conversation in computer games	0.006	0.005	179.9	1.17	0.24
Watching movies/TV programs without subtitles	-0.004	0.008	181	-0.51	0.61
Watching movies/TV programs with subtitles	0.002	0.004	180.7	0.42	0.68
Watching academic lectures	0.007	0.01	178.1	0.66	0.51
Reading instruction and/or conversation in English when playing computer games	-0.009	0.007	179.2	-1.32	0.19
Reading Internet websites	-0.01	0.007	178.7	-1.96	0.052
Reading novels	0.006	0.01	178.3	0.54	0.59
Reading magazines/newspapers/comics	0.01	0.01	180.5	1.27	0.21
Reading emails	-0.007	0.01	181	-0.71	0.48
Reading academic texts	0.0004	0.003	115.2	0.14	0.89
Reading people's posts on social networks	0.02	0.007	181	2.23	0.03 *

*** sig. at $p < 0.001$, ** sig. at $p < 0.01$, * sig. at $p < 0.05$.

4.4. The relationship between receptive collocational knowledge of high-frequency words and four factors

In answer to Research question 2, Table 5 shows that the participants' receptive knowledge of the form-meaning connection of high-frequency words had a significant, positive effect on their collocational knowledge of these words. This result indicates that as the participants' receptive knowledge of the form-meaning connection of high-frequency words increased, their collocational knowledge of these words also increased. L2 proficiency, composite exposure to L2, and length of study, however, did not make a significant contribution to the collocational knowledge of high-frequency words. Appendices 8-10 show that when the receptive collocational knowledge of high-frequency words from the 1st, 2nd, and 3rd 1,000-word levels was examined in turn, knowledge of form-meaning connection consistently contributed to learners' collocational knowledge at each frequency level. However, L2 proficiency only significantly affected the knowledge of collocations of the 1st 1,000 words, and the amount of current exposure and length of study English did not make any significant contribution to the collocational knowledge in any cases.

TABLE 5

Results of the single linear mixed-effects model analysis for collocations of high-frequency words (1st, 2nd, 3rd 1,000-word levels combined)

	Estimate	Std. Error	df	t value	Pr(> t)
(Intercept)	3.68	0.12	195.68	30.56	< 0.0001***
Composite exposure to L2	0.02	0.01	195.46	1.55	0.12
Proficiency	0.002	0.001	180.22	1.64	0.10
Length of study	-0.002	0.01	195.31	-0.32	0.75
Knowledge of high-frequency words	0.01	.001	194.59	5.98	<0.0001***

*** sig. at $p < 0.001$, ** sig. at $p < 0.01$, * sig. at $p < 0.05$. Note: Composite score refers to the sum of all types of exposure.

TABLE 6

Results of the single linear mixed-effects model analysis for the collocations of high-frequency words (individual sources of input)

	Estimate	Std. Error	df	t value	Pr(> t)
(Intercept)	3.65	0.13	180	28.78	<0.0001***
Proficiency	0.002	0.001	161.7	1.81	0.07
Length of study	-0.003	0.005	175.9	-0.51	0.61
Listening to student/teacher talks in English classes	0.009	0.009	177.3	1.01	0.31
Listening to songs	0.003	0.005	178.5	0.70	0.49
Listening to radio	-0.008	0.01	177.8	-0.59	0.56
Listening to general conversations in English outside class in person or virtually	0.01	0.01	177.7	0.73	0.47
Listening to instruction and/or conversation in computer games	0.01	0.007	176.7	1.57	0.12
Watching movies/TV programs without subtitles	-0.004	0.01	178.1	-0.35	0.73
Watching movies/TV programs with subtitles	0.002	0.006	177.1	0.41	0.69
Watching academic lectures	-0.003	0.01	179.8	-0.18	0.86
Reading the instruction and/or conversation in English when playing computer games	-0.01	0.01	176.3	-1.44	0.15
Reading Internet websites	0.008	0.01	179.7	0.81	0.42
Reading novels	0.005	0.02	179.8	0.28	0.78
Reading magazines/newspapers/comics	-0.003	0.01	177.1	-0.22	0.83
Reading emails	-0.0006	0.01	178.4	-0.05	0.96
Reading academic texts	0.004	0.004	167.3	1.13	0.26
Reading people's posts on social networks	-0.009	0.009	178.6	-1.00	0.32
Knowledge of high-frequency words	0.008	0.001	178.4	5.69	<0.0001***

*** sig. at $p < 0.001$, ** sig. at $p < 0.01$, * sig. at $p < 0.05$.

Table 6 presents the results in terms of the effects of each individual types of L2 input, L2 proficiency, length of study, and receptive knowledge of form-meaning connection of high-frequency words on the participants' receptive knowledge of collocations of high-frequency words. It shows that receptive knowledge of form-meaning connection contributed to learners' receptive collocational knowledge while other factors did not. The same patterns were found when the collocational knowledge of the 1st, 2nd, and 3rd 1,000 words was examined in turn as shown in Appendix 11, 12, and 13.

5. Discussion

By focusing specifically on high-frequency words and EFL learners from a non-cognate L1 background, the present study extends our understanding of the role of current L2 exposure, length of study, and L2 proficiency on EFL learners' vocabulary knowledge. It is also among the few studies which investigated the impact of those three factors on EFL learners' breadth and depth of vocabulary knowledge at the high-frequency levels.

5.1. The effect of EFL learners' exposure to L2 input on their receptive knowledge of form-meaning connection and collocations of high-frequency words

This study showed that the composite exposure to foreign language input did not significantly contribute to their receptive knowledge of the form-meaning connection nor collocations of high-frequency words. This result is consistent with Peters et al.'s (2019) finding of the insignificant effect of current out-of-class L2 exposure on Flemish learners' knowledge of French vocabulary. However, it contradicts most previous studies which reported a significant relationship between EFL learners' out-of-class exposure to L2 input and their vocabulary knowledge (González-Fernández & Schmitt, 2015; Muñoz, 2011; Peters, 2018; Peters et al., 2019; Puimège & Peters, 2019)

When individual sources of L2 input were examined, this study found that reading people's posts in English on social networks significantly contributed to the participants' knowledge of both high-frequency words as a whole and those from the 3rd 1000-word level, while reading Internet websites contributed to the participants' knowledge of high-frequency words from the 3rd 1,000-word level. However, none of the other examined sources of L2 input made a significant contribution to the participants' knowledge of form-meaning connection of high-frequency

words nor their collocations. It is in line with Peters' (2018) finding that browsing the Internet significantly correlated with learners' knowledge of form-meaning connection of the most frequent 5,000 words, but different from González-Fernández and Schmitt's (2015) finding that using social networking made a significant contribution to learners' collocational knowledge of the most frequent 5,000 words. This also contrasts previous research which reported the significant effect of watching movies/TV programs (González Fernández, 2021; Peters, 2018; Puimège & Peters, 2019), reading magazines, reading books (Peters, 2018), playing games (Peters, 2018; Puimège & Peters, 2019) on EFL learners' knowledge of the most frequent 5,000 words, and the significant effect of reading, watching movies/TV, and immersion in English-speaking countries on their collocational knowledge of these words (González Fernández & Schmitt, 2015).

There are several possible reasons for the different findings of this study and previous studies. First, as shown in Figure 3, the most frequent type of out-of-class exposure to L2 input for the participants in the present study was reading academic texts. Academic texts may better facilitate the learning of specialized vocabulary than general vocabulary. Meanwhile, empirical studies have provided solid evidence that other types of input such as TV programs (e.g., Peters & Webb, 2018) and games (e.g., De Wilde et al., 2019; Sylvén & Sundqvist, 2012) could positively facilitate the learning of general vocabulary. Yet the participants' exposure to these L2 sources was very limited. Second, previous research examined learners' knowledge of form, meaning, and collocations of the most frequent 5,000 words, 6,000 words, and 10,000 words while this study investigated knowledge of form, meaning, and collocations of high-frequency words (the most frequent 3,000 words). The difference in the frequency of the examined vocabulary might also account for the conflicting findings. Third, previous research either examined learners from cognate L1 backgrounds or did not distinguish those from cognate and non-cognate L1 backgrounds whereas this study focused solely on learners from a non-cognate L1 background. Due to linguistic similarities between cognate languages, it is easier for learners from cognate L1 backgrounds to learn vocabulary than learners from non-cognate L1 backgrounds (De Wilde et al., 2021; Nation, 2022; Peters, 2020).

Overall, even though the non-cognate EFL learners at the postgraduate level were regularly exposed to English, there is a lack of significant relationship between their current L2 exposure

and vocabulary knowledge. This finding further support González Fernández and Schmitt's (2015) point that it may not be the amount of input that matters but the quality of engagement with the input does.

5.2. The effect of EFL learners' length of study on their receptive knowledge of form-meaning connection and collocations of high-frequency words

The current study revealed that the length of studying English did not significantly predict learners' receptive knowledge of form-meaning connection and collocations of high-frequency words. This finding contrasts earlier studies, which reported that the length of instruction positively correlated with learners' vocabulary knowledge (González-Fernández & Schmitt, 2015; Muñoz, 2011; Peters, 2018; Peters et al., 2019). The conflicting findings could be explained in three ways. First, as mentioned, participants in previous research were mainly cognate L1 learners, and therefore may have more advantages when learning English vocabulary than the participants in the present study who were from a non-cognate L1 background. Second, the participants in the previous studies were mainly undergraduates or secondary school students, while the participants in the present study were postgraduate students. Compared with secondary school students and undergraduates, postgraduates probably had a longer period of English instruction and usually came from a wider range of (thus more mixed) educational contexts (Feak, 2016). In such context, other factors such as intensity of instruction rather than length of instruction might be a more important factor which contributed to postgraduates' vocabulary knowledge. Previous research has shown that learners who studied English for a shorter time but with intense instruction were able to perform better than their peers with the same amount of instruction distributed over a longer period of time (Collins et al., 1999). This might be the case for some postgraduates in the present study. Last, as suggested by an anonymous reviewer, one possible reason why length of study was not a significant predictor might be that the group examined in the present study was quite homogeneous in this respect with the standard deviation being 2.10.

5.3. The effect of L2 proficiency on EFL learners' receptive knowledge of form-meaning connection and collocations of high-frequency words

This study found that L2 proficiency significantly and positively affected the participants' receptive knowledge of form-meaning connection of high-frequency words. This finding is

consistent with that of Gui (2015) who found who found a positive relationship between Chinese EFL learners' vocabulary breadth and L2 proficiency as represented by their scores on the College English Test Band 4. This is probably not surprising as our participants were at a proficiency level of just passing the College English Test Band 4. One possible reason for the significant effect of L2 proficiency on learners' knowledge of form-meaning connection of high-frequency words might be the close relationship between EFL learners' knowledge of form-meaning connection and their L2 proficiency. Research has shown that knowledge of form and meaning is a positive predictor of EFL learners' overall L2 proficiency (e.g., Miralpeix & Muñoz, 2018; Schmitt, 2014), even at the basic level of proficiency (Schmitt & Schmitt, 2014). Therefore, it is not surprising that the reverse effect also occurred.

Unlike the case of form-meaning connection, L2 proficiency only significantly affected the participants' receptive collocational knowledge of the 1st 1000 words. This finding is partially in line with González Fernández and Schmitt (2015), who found a strong positive correlation between learners' self-rated proficiency and their collocational knowledge of the most frequent 5,000 words. There were several crucial differences between the present study and González Fernández and Schmitt's (2015) study, which made further comparison difficult. First, while González Fernández and Schmitt asked their participants to self-rate their proficiency, the current study used the English scores of the National Postgraduate Entrance Examination, which provided a more precise idea of the participants' language proficiency. Second, the dimensions of the collocations measured in the two studies were different. González Fernández and Schmitt focused on productive knowledge while the present study looked at receptive knowledge. Third, González Fernández and Schmitt's participants came from a cognate L1 background (Spanish) while those in the present study were from a non-cognate L1 background (Chinese). Last, González Fernández and Schmitt examined collocational knowledge of the most frequent 5,000 words while the present study examined collocation knowledge of high-frequency words.

There are two possible reasons why L2 proficiency only made a significant contribution to the learning of collocations of the first 1,000 words. First, the participants' insufficient knowledge of the form-meaning connection of the 2nd and 3rd 1,000 words might be a reason why L2 proficiency did not significantly affect the collocational knowledge of words at these frequency levels. Results of the UVLT showed that as a whole the participants only nearly reached the

mastery level of the 1st 1,000 words, and had not mastered the 2nd and 3rd 1,000 words in terms of form-meaning connection. Second, the learners' even more limited knowledge of collocations of words from the 2nd and 3rd 1,000 levels (less than 50%) might be another reason for the insignificant relationship between their L2 proficiency and knowledge of collocations of these words.

5.4. Comparison with Nguyen and Webb (2017) for EFL learners' receptive knowledge of form-meaning connection and collocations of high-frequency words, and their relationship

Similar to Nguyen and Webb (2017), our study used the UVLT to measure the participants' knowledge of form-meaning connection and the RCT to measure their collocational knowledge of these words (including verb-noun and adjective-noun collocations). However, it expands on Nguyen and Webb's study in several ways. First, Nguyen and Webb examined the effect of word-related factors on collocational knowledge of high-frequency words, this study examined the effect of learner-related factors (current L2 exposure, length of study, and L2 proficiency) on knowledge of both form-meaning connection and collocations of high-frequency words. As discussed in sections 5.1 to 5.3, looking at the issue from these perspectives provides further insight into the impact of various factors on the breadth and depth of knowledge of high-frequency words. Second, Nguyen and Webb's study was conducted with Vietnamese English majored undergraduate students. By examining Chinese non-English majored postgraduate students, this study extends Nguyen and Webb's (2017) finding by suggesting that the positive correlation between EFL learners' knowledge of form-meaning connection and collocations of high-frequency words also holds true to non-English major postgraduate students from other non-cognate L1 backgrounds.

The positive relationship between non-cognate EFL learners' knowledge of form-meaning connection of high-frequency words and collocational knowledge of these words indicates that as these learners' vocabulary breadth improves, their depth of vocabulary knowledge also tends to improve. This finding thus corroborates the view that vocabulary breadth and depth are two closely related dimensions in the development of L2 vocabulary knowledge (Akbarian, 2010; González Fernández & Schmitt, 2020; Milton, 2009; Schmitt, 2014). In addition, this study found that the postgraduate students mastered 96.47%, 82.80%, and 71.64% of the words from the 1st 1,000, 2nd and 3rd 1,000-word levels in terms of form-meaning connection. However, they

only knew nearly 58%, 46%, and 35% of the collocations of these words from each level respectively. These findings indicated that Chinese EFL learners had insufficient knowledge of form-meaning connection and collocations of high-frequency words. Such findings are comparable to that of Nguyen and Webb (2017) for English majors at the undergraduate level. It means that EFL learners' collocational knowledge of high-frequency words appeared not to be as great as their form-meaning connection knowledge of these words. These findings added solid evidence supporting Schmitt (2014), who concluded from previous empirical studies that L2 learners with a vocabulary breadth of 3,000 to 10,000 word families can have a good knowledge of collocations at the recognition level, whereas those with a vocabulary size of lower than 3,000 word families may have limited knowledge of collocations. Together, these studies lent support to the claim that learners' collocational knowledge is likely to lag behind their knowledge of form-meaning connection (Akbarian, 2010; Bahns & Eldaw, 1993; González Fernández, 2021; Irujo, 1993). There are two possible reasons for EFL learners' lack of collocational knowledge. First, vocabulary teaching and learning in China tended to place more emphasis on form and meaning rather than collocations (see Gu, 2003 for an example). Therefore, the learners were likely to focus on expanding their vocabulary breadth rather than developing vocabulary depth. Second, the participants' insufficient knowledge of form-meaning connection of words would then have a negative impact on their development of the collocational knowledge of these words. The findings of the present study thus call for attention to the breadth and depth of vocabulary knowledge at high-frequency levels in EFL teaching and learning.

5.5. EFL learners' current exposure to L2 input

The present study showed that reading academic texts is the most frequent activity, followed by listening to teachers/students' speaking in English language classes, listening to songs, and watching movies/TV programs with subtitles. These findings are consistent with much of previous research which indicated that songs, movies/TV programs, computer games, internet and social networks were among the most frequent types of L2 input (e.g., González Fernández 2021; Peters et al., 2019). However, it expands on previous research in several ways. First, reading academic texts was found to be the most frequent kind of input in the present study. This finding contrasts previous studies which showed that reading was the least frequent type of L2 input (Peters, 2018; Peters et al., 2019; Puimège & Peters, 2019). This difference may be because the participants in the present study were postgraduates while those from the previous studies

were undergraduates or even secondary school and primary school students. Previous research indicated that learners at higher educational levels were likely to do more reading than those at lower education levels (Peters et al., 2019) to accommodate to the needs of their study. This study thus adds further evidence that learners' exposure to L2 input indeed varied according to learners' educational levels and highlights the value of investigating postgraduate students in this study. Second, listening to teachers/students' speech in English language class was found to be the second most frequent kind of input. This finding is interesting. Teacher talk is a fundamental type of oral input in second language learning (Van Patten, 2003). Both corpus research (e.g., Coxhead, 2017; Horst, 2010; Macaro & Tian, 2015) and experimental research (Jin & Webb, 2020) have suggested that it is a potential source for incidental L2 vocabulary learning. Yet, previous research on the input that EFL learners were exposed to only focused on out-of-class input. Therefore, the finding related to listening to teachers/students' speaking in English language classes in this study is a novel contribution to research on EFL learners' input. Taken together, this study provides us with a relatively comprehensive understanding of the amount of L2 input that non-cognate L1 EFL learners were exposed to.

6. Pedagogical implications

This study has several pedagogical implications. First, the EFL learners, even at the postgraduate level, should consistently consolidate and expand their knowledge of high-frequency words. This study found that despite 10 years or more of studying English, the participants had insufficient knowledge of these lexical items. Therefore, more attention should be paid to the breadth and depth of high-frequency word knowledge. This could be done by raising students' awareness of the importance of high-frequency words for communication, using the UVLT and RCT to diagnose students' knowledge of high-frequency words and their collocations, and organize learning activities across Nation's (2007) Four Strands for students to repeatedly meet and use high-frequency words in a meaningful way (see Dang, 2018 for further information).

In terms of out-of-class activities, this study showed that although the participants had initial exposure with different types of L2 input, their frequency of exposure to those types of input, which were proved to be effective, might be low. Thus, teachers should encourage their students to make good use of such sources, for example, viewing TV programs. Moreover, the teachers

should help the learners improve their quality of engagement with the language input by designing related-activities such as story retelling and role play to be presented in class. In terms of in-class activities, the fact that listening to teachers/students' speech in English language class is a frequent kind of L2 input suggests that to optimize students' exposure to high-frequency words and their collocations, teachers could try to incorporate these lexical items in their talks and organize activities for students to frequently encounter and use these items in meaningful ways.

Second, this study showed that learners' L2 proficiency is an important predictor of their vocabulary knowledge. Therefore, apart from activities which intentionally help EFL learners to learn vocabulary, it is equally important to promote activities which help to improve learners' overall language proficiency (e.g., interactive reading, writing with feedback). These activities also provide learners with opportunities for incidentally vocabulary learning.

Third, this study revealed that length of study did not significantly affect the postgraduates' vocabulary knowledge in the EFL contexts. Therefore, EFL teachers, material writers, and course designers should not assume that the more senior students, the better their knowledge of high-frequency words and their collocations. Instead, at the beginning of a language program, teachers should diagnose their learners' knowledge of high-frequency words and their collocations to ensure that they have sufficient knowledge of these words, before further instruction of words at lower frequency levels is provided.

Last but not least, the significant effect of knowledge of form-meaning connection of high-frequency words on their collocational knowledge highlights the importance of developing EFL learners' knowledge of form-meaning connection of high-frequency words as the starting point of vocabulary learning. In addition, given that the learners had limited collocational knowledge of high-frequency words, it is suggested that more deliberate attention should be given to collocations in learning and teaching. For instance, the teachers could raise awareness about the importance of collocations in fluent comprehension and production, as well as design activities to help their students notice, understand, and use collocations in texts (e.g., Nation, 2022; Webb & Nation, 2017). They could also use typographic enhancement to help the learners notice collocations (e.g., Boers et al., 2017; Toomer & Elgort, 2019). Meanwhile, the material writers

can make the collocations of high-frequency vocabulary more visible, for example by glossing, in the learning materials.

7. Conclusion

The present study is among the few studies which examined the effect of current L2 exposure, length of study, and L2 proficiency on non-cognate EFL learners' receptive knowledge of form-meaning connection and collocations of high-frequency words. It revealed that current L2 exposure and length of study did not seem to have an impact on learners' knowledge of form-meaning connection of high-frequency words nor their collocations. However, L2 proficiency significantly contributed to these learners' receptive knowledge of form-meaning connection and collocations of high-frequency words from the 1st 1,000-word level. Moreover, receptive knowledge of form-meaning connection of high-frequency words also contributed to the collocational knowledge of these words. Overall, this study extended our understanding towards the role of current L2 exposure, length of study, and L2 proficiency on two dimensions of vocabulary knowledge (i.e., vocabulary breadth and depth), which is an under-researched area. It also provided us with some important implications into the teaching and learning of high-frequency vocabulary.

This study has several limitations. First, the participants' vocabulary knowledge was only measured at the receptive level, productive knowledge may provide further insights into learners' vocabulary knowledge. Second, following previous research (e.g., González Fernández & Schmitt, 2015; Peters, 2018; Peters et al., 2019), the self-reported questionnaire format was used to collect information about current L2 exposure, length of study, and L2 proficiency. To better reflect students' experience and language proficiency, future research may use other methods (e.g., case studies) to triangulate and supplement the information from self-reported questionnaires and collect actual language test scores rather than using students' self-reported scores. Third, future research which investigates the types of input (e.g., amount of time visiting English-speaking countries) more fully would be valuable. Last, longitudinal research employing other methods (e.g., experiment, ethnography) would provide novel insights into the effects of L2 exposure on EFL learners' vocabulary knowledge.

Notes

¹ Following Nguyen and Webb (2017), the present study defines collocation as pairs of words which co-occur more frequently than by chance.

² According to Dyen et al. (1992) and Schepens et al. (2011), English shares many cognates with both Germanic (e.g., Dutch, German) and Romance (e.g., Spanish, French) languages. Learners from such L1 backgrounds are thus called cognate L1 learners in the present study.

³ Test samples can be found in this document.

Ministry of Education Examination Center (2018). *Unified National Graduate Entrance Examination of 2019 for Non-English majors: Test syllabus of English (I)*. Higher Education Press.

Acknowledgements: We would like to thank the three anonymous reviewers. Their advice has significantly strengthened this manuscript. Our sincere thanks also go to all the participants for taking part in this study, and all the teachers for helping with data collection. In addition, we would like to express our special thanks to the Editor Professor Hossein Nassaji for support throughout the review process.

References

- Akbarian, I. h. (2010). The relationship between vocabulary size and depth for ESP/EAP learners. *System*, 38(3), 391–401. <https://doi.org/10.1016/j.system.2010.06.013>
- Alsaif, A., & Milton, J. (2012). Vocabulary input from school textbooks as a potential contributor to the small vocabulary uptake gained by English as a foreign language learners in Saudi Arabia. *The Language Learning Journal*, 40(1), 21–33. <https://doi.org/10.1080/09571736.2012.658221>
- Bahns, J., & Eldaw, M. (1993). Should we teach EFL students collocations. *System*, 21, 101–114. [https://doi.org/10.1016/0346-251X\(93\)90010-E](https://doi.org/10.1016/0346-251X(93)90010-E)
- Boers, F., Demecheleer, M., He, L., Deconinck, J., Stengers, H., & Eyckmans, J. (2017). Typographic enhancement of multiword units in second language text. *International Journal of Applied Linguistics*, 27(2), 448–469. <https://doi.org/10.1111/ijal.12141>
- Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education* (8th ed.). Routledge.
- Collins, L., Halter, R. H., Lightbown, P. M., & Spada, N. (1999). Time and the distribution of time in L2 instruction. *TESOL Quarterly*, 33, 655–680.
- Coxhead, A. (2017). Academic vocabulary in teacher talk: challenges and opportunities for pedagogy. *Oslo Studies in Language*, 9(3), 29–44.
- Cunnings, I. (2012). An overview of mixed-effects statistical models for second language researchers. *Second Language Research*, 28(3), 369–382.
- Dang, Y. (2018). Teaching the most useful words of English. In J. I. Liantas, M. DelliCarpini, S. Abrar-ul-Hassan, N. J. Anderson, D. D. Belcher, & C. Broady (Eds.), *The TESOL Encyclopedia of English Language Teaching* (pp. 1–7). Wiley-Blackwell.
- Dang, T. N. Y. (2020a). High-frequency words in academic spoken English: corpora and learners. *ELT Journal*, 74(2), 146–155. <https://doi.org/10.1093/elt/ccz057>
- Dang, T. N. Y. (2020b). Vietnamese non-English major EFL university students' receptive knowledge of the most frequent English words. *VNU Journal of Foreign Studies*, 36(3). <https://doi.org/https://doi.org/10.25073/2525-2445/vnufs.4553>
- Dang, T. N. Y., & Webb, S. (2020). Vocabulary instruction and the good language teachers. In Z. T. C. Griffiths, & A. Brown (Ed.), *Lessons from good language teachers* (pp. 203–218). Cambridge University Press.

- Dang, T. N. Y., Webb, S., & Coxhead, A. (2020). Evaluating lists of high-frequency words: Teachers' and learners' perspectives. *Language Teaching Research*, 1–25.
- De Wilde, V., Brysbaert, M., & Eyckmans, J. (2019). Learning English through out-of-school exposure: Which levels of language proficiency are attained and which types of input are important? *Bilingualism: Language and Cognition*, 23(1), 171–185.
<https://doi.org/10.1017/s1366728918001062>
- De Wilde, V., Brysbaert, M., & Eyckmans, J. (2021). Formal versus informal L2 learning. *Studies in Second Language Acquisition*, 1–25.
<https://doi.org/10.1017/s0272263121000097>
- Dörnyei, Z. (2007). *Research Methods in Applied Linguistics: Quantitative, Qualitative and Mixed Methodologies*. Oxford University Press.
- Dyen, I., Kruskal, J.B. & Black, P. (1992). An Indoeuropean classification: A lexicostatistical experiment. *Transactions of the American Philosophical Society*, 82, iii–132.
- Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, 24(2), 143–188.
- Feak, C. (2016). EAP support for post-graduate students. In K. Hyland & P. Shaw (Eds.), *The Routledge handbook of English for academic purposes* (pp. 489–501). Routledge.
- González-Fernández, B. (2021). Vocabulary knowledge: Lexical depth and its relationship with out-of-class exposure. In P. Szudarski & S. Barclay (Eds.), *Vocabulary Theory, Patterning and Teaching* (pp. 21–47). Multilingual Matters.
- González Fernández, B., & Schmitt, N. (2015). How much collocation knowledge do L2 learners have? *ITL - International Journal of Applied Linguistics*, 166(1), 94–126.
<https://doi.org/10.1075/itl.166.1.03fer>
- González-fernández, B., & Schmitt, N. (2020). Word knowledge: Exploring the relationships and order of acquisition of vocabulary knowledge components. *Applied Linguistics*, 41(4), 481–505. <https://doi.org/10.1093/applin/amy057>
- Gu, Y. P. (2003). Fine brush and freehand: The vocabulary-learning art of two successful Chinese learners. *TESOL Quarterly*, 37(1), 73–104.
- Gui, L. (2015). Predictability of vocabulary size on learners' EFL proficiency: Taking VST, CET4 and CET6 as instruments. *Studies in Literature and Language*, 10 (3), 18–23.

- Gyllstad, H. (2009). Designing and evaluating tests of receptive collocation knowledge: COLLEX and COLLOMATCH. In A. Barfield & H. Gyllstad (Eds.), *Researching collocations in another language: Multiple interpretations* (pp. 153–170). Palgrave Macmillan.
- Henriksen, B., & Danelund, L. (2015). Studies of Danish L2 learners' vocabulary knowledge and the lexical richness of their written production in English. In K. D. P. Pietilä, & R. Pipalov a (Ed.), *Lexical issues in L2 writing*. Cambridge Scholars Publishing.
- Horst, M. (2010). How well does teacher talk support incidental vocabulary acquisition? *Reading in a foreign language*, 22, 161–180.
- Irujo, S. (1993). Steering clear: Avoidance in the production of idioms. . *International Review of Applied Linguistics in Language Teaching*, 31, 205–219.
<https://doi.org/10.1515/iral.1993.31.3.205>
- Jin, Z., & Webb, S. (2020). Incidental vocabulary learning through listening to teacher talk. *The Modern Language Journal*, 104(3), 550–566. <https://doi.org/10.1111/modl.12661>
- Jordan, G., & Gray, H. (2019). We need to talk about coursebooks. *ELT Journal*, 73(4), 438–446. <https://doi.org/10.1093/elt/ccz038>
- Laufer, B., & Ravenhorst-Kalovski, G. C. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language*, 22(1), 15–30.
- Laufer, B., & Waldman, T. (2011). Verb-noun collocations in second language writing: A corpus analysis of learners' English. *Language Learning*, 61, 647–672.
- Lu, C., & Dang, T. N. Y. (2022). Vocabulary in EAP learning materials: What can we learn from teachers, learners, and corpora? *System*, 106.
<https://doi.org/10.1016/j.system.2022.102791>
- Macaro, E., & Tian, L. (2015). Exploring teachers' oral explanations of new English lexical items in a Chinese university: Comparisons with dictionary information. *System*, 52, 78–90. <https://doi.org/10.1016/j.system.2015.05.002>
- Matthews, J., & Cheng, J. (2015). Recognition of high frequency words from speech as a predictor of L2 listening comprehension. *System*, 52, 1–13.
<https://doi.org/10.1016/j.system.2015.04.015>
- Milton, J. (2009). *Measuring second language vocabulary acquisition*. Multilingual Matters.

- Ministry of Education Examination Center (2018). *Unified National Graduate Entrance Examination of 2019 for Non-English majors: Test syllabus of English (I)*. Higher Education Press.
- Miralpeix, I., & Muñoz, C. (2018). Receptive vocabulary size and its relationship to EFL language skills. *International Review of Applied Linguistics in Language Teaching*, 56(1), 1–24. <https://doi.org/10.1515/iral-2017-0016>
- Muñoz, C. (2011). Input and long-term effects of starting age in foreign language learning. *IRAL - International Review of Applied Linguistics in Language Teaching*, 49(2), 113–133. <https://doi.org/10.1515/iral.2011.006>
- Nation, I. S. P. (1983). Testing and teaching vocabulary. *Guidelines*, 5(1), 12–25.
- Nation, I. S. P. (2007). The four strands. *Innovation in Language Learning and Teaching*, 1(1), 1–12.
- Nation, I. S. P. (2012). *The BNC/COCA word family lists*. In <http://www.victoria.ac.nz/lals/about/staff/paul-nation>.
- Nation, I. S. P. (2022). *Learning vocabulary in another language* (3rd ed.). Cambridge University Press.
- Nguyen, C.-D. (2020). Lexical features of reading passages in English-language textbooks for Vietnamese high-school students: Do they foster both content and vocabulary Gain? *RELC Journal*, 1–14. <https://doi.org/10.1177/0033688219895045>.
- Nguyen, T. M. H., & Webb, S. (2017). Examining second language receptive knowledge of collocation and factors that affect learning. *Language Teaching Research*, 21(3), 298–320.
- Olmos, C. (2009). An assessment of the vocabulary knowledge of students in the final year of secondary education. Is their vocabulary extensive enough? *International Journal of English Studies*(Special Issue), 73–90.
- Peters, E. (2018). The effect of out-of-class exposure to English language media on learners' vocabulary knowledge. *ITL - International Journal of Applied Linguistics*, 169(1), 142–168. <https://doi.org/10.1075/itl.00010.pet>
- Peters, E., Noreillie, A. S., Heylen, K., Bulté, B., & Desmet, P. (2019). The impact of instruction and out-of-school exposure to foreign language input on learners' vocabulary knowledge

- in two languages. *Language Learning*, 69(3), 747–782.
<https://doi.org/10.1111/lang.12351>
- Peters, E., & Webb, S. (2018). Incidental vocabulary acquisition through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition*, 40(3), 551–577.
<https://doi.org/10.1017/s0272263117000407>
- Puimège, E., & Peters, E. (2019). Learners’ English vocabulary knowledge prior to formal instruction: The role of learner-related and word-related variables. *Language Learning*, 69(4), 943–977. <https://doi.org/10.1111/lang.12364>
- Read, J., & Dang, T. N. Y. (2022). Measuring depth of academic vocabulary knowledge. *Language Teaching Research*. <https://doi.org/10.1177/13621688221105913>
- Schepens, J. O. B., Dijkstra, T. O. N., & Grootjen, F. (2011). Distributions of cognates in Europe as based on Levenshtein distance. *Bilingualism: Language and Cognition*, 15(1), 157–166. <https://doi.org/10.1017/s1366728910000623>
- Schmitt, N. (2014). Size and depth of vocabulary knowledge: What the research shows. *Language Learning*, 64 (4), 913–941.
- Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *The Modern Language Journal*, 95(1), 26–43.
- Schmitt, N., & Schmitt, D. (2014). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching*, 47(04), 484–503.
- Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, 18(1), 55–88.
- Siyanova-Chanturia, A., & Pellicer-Sánchez, A. (2018). *Understanding formulaic language: A second language acquisition perspective*. Routledge.
- Sun, Y., & Dang, T. N. Y. (2020). Vocabulary in high-school EFL textbooks: Texts and learner knowledge. *System*, 93, 1–13. <https://doi.org/10.1016/j.system.2020.102279>
- Sylvén, L. K., & Sundqvist, P. (2012). Gaming as extramural English L2 learning and L2 proficiency among young learners. *ReCALL*, 24(3), 302–321.
<https://doi.org/10.1017/s095834401200016x>

- Toomer, M., & Elgort, I. (2019). The development of implicit and explicit knowledge of collocations: A conceptual replication and extension of Sonbul and Schmitt (2013). *Language Learning*, 69(2), 405–439. <https://doi.org/10.1111/lang.12335>
- Van Patten, B. (2003). *From input to output: A teacher's guide to second language acquisition*. McGraw–Hill.
- Webb, S., & Nation, I. S. P. (2017). *How vocabulary is learned*. Oxford University Press.
- Webb, S., Sasao, Y., & Ballance, O. (2017). The updated Vocabulary Levels Test. *ITL - International Journal of Applied Linguistics*, 168(1), 33–69. <https://doi.org/10.1075/itl.168.1.02web>