Relationships Between Lexical Coverage, Learner Knowledge, and Teacher Perceptions of the Usefulness of High-frequency Words

Abstract

Recently researchers have proposed using information from teachers and learners to supplement the information from corpora in the selection of the most useful words for foreign language learners. Yet the extent to which these data sets correlate to one another is unclear. This study explicitly investigated the relationships between (a) the lexical coverage in 18 corpora, (b) the knowledge of 275 Vietnamese EFL learners, and (c) the perceptions of 78 EFL teachers of the usefulness of 973 high-frequency words. The correlations between lexical coverage and the other two factors were significant but small, and the correlations between learner vocabulary knowledge and teacher perceptions were large. Teacher perceptions better predicted learner knowledge than lexical coverage. This study confirms the value of lexical coverage as a key criterion to select words for foreign language learners, and also highlights the importance of teachers' perceptions of usefulness on ultimate vocabulary learning in EFL contexts.

Key words*: learner vocabulary knowledge; teacher cognition; lexical coverage; high-frequency words; corpus linguistics*

1. Introduction

There are a large number of words in the target language that foreign language learners need to know, but the number of words that they can acquire per year is fairly modest (Milton, 2009; Webb & Chang, 2012). To ensure that their learning time is well-spent, these learners should start their vocabulary learning from the words that are most useful for them (Laufer & Nation, 2012). Corpus linguistic research offers an innovative approach toward identifying these words. Based on the analysis of vocabulary in corpora representing language use at large (e.g., general conversation, academic texts, lectures, newspapers, novels, and movies), researchers have found that while a large number of words (23,000 or more words) (e.g., *gloat, petunia*) only occur several times or once in the target language, a smaller number of words (around 2,000 words) (e.g., *good, know)* occur very frequently and make up a major proportion of texts in a wide range of discourse types (70%-90%) (Dang & Webb, 2020; Nation, 2006). The words in the former group are low-frequency words while those in the latter are high-frequency words. From the cost-effective perspective, high-frequency words are more useful for foreign language learners to know than low-frequency words, because learners only need to study a much smaller number of words but are able to cope with a larger percentage of words in various texts. As comprehension increases along with the growth in the percentage of known words in texts (Schmitt, Jiang, & Grabe, 2011), knowledge of high-frequency words would provide foreign language learners with a better chance to comprehend language which is used in a wide range of situations, particularly compared to knowledge of low-frequency words. Given the value of high-frequency words, frequency-based criteria such as lexical coverage in corpora representing language use at large (i.e., the percentage of words in a corpus covered by a word or items from a particular word list) have been widely used to identify the most useful words for foreign language learners, especially

those learning English as a foreign language (EFL) (e.g., Brezina & Gablasova, 2015; Nation, 2006; 2012).

Despite the great value of high-frequency words for foreign language learners and the innovative method that corpus linguistics has offered to identify these words, practitioners (e.g., Stein, 2017) raise the concern that teachers and learners may not see clearly the value of corpus-based high-frequency word lists to their teaching because they may think that some items in these lists are not very relevant to foreign language learners. This claim is supported by Dang and Webb's (2020) survey with EFL teachers which showed that corpus-based word lists were the least popular source for vocabulary instruction. Moreover, researchers (Nation, 2016; Laufer & Nation, 2012) have pointed out that frequency-based information from corpora (e.g., lexical coverage) may not be the only factor that determines the usefulness of a word for learners for several reasons. As corpora represent language use at large, there is likely to be a lot of variation between the language that makes up a corpus and the language encountered by foreign language learners in a particular learning context (Brysbaert, Keuleers, & Mandera, 2020). Other factors (e.g., tests, materials) may also influence decisions on the usefulness of words.

Therefore, recently word lists researchers have proposed using information from other sources to supplement the information from corpora in the selection of the most useful words for foreign language learners. Several studies sought for opinions from teachers (Dang, Webb, & Coxhead, 2022; He & Godfroid, 2018), and others measured learners' vocabulary knowledge (Dang *et al*., 2022; Brysbaert *et al*., 2020). While each set of information from corpora, teachers, and learners brings interesting insights into the nature of the most useful words for foreign language learners, one question that arises is to what extent these data sets correlate to one another. No studies have explicitly addressed this question. To address this gap, the present study

quantitatively examined the relationships between (a) lexical coverage in corpora representing language use at large, (b) EFL learner vocabulary knowledge, and (c) EFL teacher perceptions of the usefulness of each high-frequency word. By bringing together different kinds of data on high-frequency words, this study enriches our understanding of the relative value of the key criteria used in the selection of the most useful words for foreign language learners and provide useful implications for the development and implementation of corpus-based high-frequency word lists in foreign language contexts.

2.   Literature review

2.1. Relationship between lexical coverage in corpora and learner knowledge of high-frequency words

There does not appear to be any research examining the relationship between lexical coverage in corpora representing language use at large and foreign language learner knowledge of each high-frequency word. It is important to examine the relationship between these two factors because it indicates the extent to which foreign language learners know the words that are essential for them to comprehend and communicate in various situations (e.g., general conversation, academic texts, lectures, newspapers, novels, and movies). Research indicates that the relationship between lexical coverage and learners' vocabulary knowledge is likely to be very high because words that provide higher lexical coverage in corpora are more frequent and words tend to be learned according to their frequency in the language (e.g., Dóczi & Kormos, 2016; Edwards & Collins, 2011; Horst & Collins, 2006; Schmitt, Schmitt, & Clapham, 2001). However, the extent to which this is true for words that are close in frequency has never been investigated.

4

There are three reasons why the relationship between these two variables might be different for words from the same frequency level. First, although words tend to be learned by frequency, knowledge of words varies from learner to learner. The words known by one learner will be different from the words known by another learner. Second, the fact that words that are similar in frequency are not learned may be influenced by a range of other factors such as lexical difficulty, cognancy, the way that they are taught, and the contexts in which they are encountered rather than their frequencies (Brysbaert *et al.*, 2020; Laufer, 1990, 1997). Third, items in corpus-based word lists have been mainly selected based on their frequency in corpora that aim to represent language use at large. Such language may not correspond to the actual frequency with which learners are exposed to specific words in their specific learning context. The influence of frequency and in turn lexical coverage will largely be affected by the kind and amount of input that learners encounter (Webb & Nation, 2017). In many foreign language contexts, the amount of input is fairly limited (Muñoz 2008; Webb & Nation, 2017), and mainly restricted to language classrooms (Laufer, 2001; 2003), especially textbooks and teacher talk (Horst, 2010; Richards, 2010). However, research has shown that textbooks provide little opportunities for learners to deliberately and incidentally learn high-frequency words (e.g., Sun & Dang, 2020). Meanwhile, both corpus-driven and experimental studies have revealed that although high-frequency words accounted for around 95% of the words in teacher talk (Coxhead, 2017; Horst, 2010) and learners could incidentally learn vocabulary from listening to teacher talk (Jin & Webb, 2020), the learning gains are relatively small (Jin & Webb, 2020). Given the little opportunities for learners to learn vocabulary through textbooks and teacher talk, frequency in the corpora representing language use at large, from which high-frequency words were selected, may have a small effect on the words that foreign language learners learn. Investigating the relationship

between lexical coverage in corpora and foreign language learner knowledge will provide a more accurate indication of the extent to which the words considered as being useful from the frequency-based perspective are actually learned by foreign language learners.

2.2. *Relationship between frequency in corpora and teacher perceptions of usefulness of high-frequency words*

To the best of our knowledge, only Ellis, Simpson-Vlach, and Maynard (2008) and He and Godfroid (2018) have examined the relationship between frequency in corpora and teacher perceptions of word usefulness. Both studies found significant but small (*r=.26*) (Ellis *et al*., 2008) or medium (*r=.47*) correlations (He & Godfroid, 2018) between the two variables. As lexical coverage in corpora is a frequency-based criterion, it can be inferred from these studies that lexical coverage in corpora is only one factor that determines the value of a lexical item for language learners, and teacher perceptions can provide further insights into the value of the item. However, it should be noted that Ellis *et al*. (2008) and He and Godfroid (2018) conducted their studies in English as a second language contexts and focused on academic vocabulary rather than high-frequency words and only investigated the perceptions of teachers in a specific institution.

According to Borg (2006), teacher perceptions or cognition (i.e., what teachers think, believe, and know) are influenced by various factors such as their prior language learning experience, teacher education, classroom practice, and contextual factors. Research exploring the relationship between the lexical coverage in corpora and teacher perceptions of the usefulness of each high-frequency word would provide an ecologically valid assessment of the validity of corpus-based high-frequency word lists. Moreover, while there is an assumption in word list research that high-frequency items in corpora are useful for learners in all contexts (e.g., Dang, 2018; Nation, 2013; Schmitt, 2010; Webb & Nation, 2017), examining the perceptions of high-

frequency words with experienced teachers from a range of EFL contexts may indicate the extent to which this assumption holds true. Exploring teachers' perceptions of word usefulness and juxtaposing them with insights from corpus-based research into lexical coverage and word lists would shed light on the role of teachers in promoting vocabulary.

2.3.   *Relationship between teacher perceptions of word usefulness and learner knowledge of high-frequency words*

There does not appear to be any research examining the relationship between teacher perceptions of word usefulness and foreign language learner knowledge of high-frequency words. This is surprising because the relationship between the two factors can show the extent to which teacher perceptions of word usefulness may influence what students actually learn.

Vocabulary can be learned both explicitly and incidentally. Explicit learning means vocabulary is the main focus of learning activities (e.g., word cards, crosswords) while incidental learning means vocabulary learning is a by-product of other activities (e.g., listening to songs, reading novels, watching television programs) (Webb & Nation, 2017). Research on second language acquisition suggests that second language learners may acquire words from natural exposure to the target language; that is, words may be learned incidentally through repeated encounters in input (Ellis, 2002). Thus, more frequent words may likely be learned before less frequent words. As a result, corpus-based word lists are created with the aim of guiding students to learn the words which have high frequency in the target language input. However, in many foreign language contexts, learners have very limited exposure to the target language (Muñoz, 2008). Therefore, teachers may play an important role in vocabulary learning as they provide learners opportunities for both incidental and explicit learning (Laufer, 2001, 2003). Yet, no studies have

been conducted to examine which source, natural exposure to the target language or teachers, better contributes to the learning of high-frequency words.

3. **Present study**

The present study is one attempt to bring together different kinds of data on high-frequency vocabulary. This is the first study that has quantitatively examined the relationship between (a) lexical coverage from corpora representing language use at large, (b) foreign language learner vocabulary knowledge, and (c) foreign language teacher perceptions of the usefulness of each high-frequency word. Focusing specifically on English as a foreign language, this study used a large number of items from current high-frequency word list as the target words and investigated the lexical coverage of these words in a large number of English corpora with a great degree of diversity, the perceptions of EFL teachers from a wide range of contexts, and the vocabulary knowledge of EFL learners with different proficiency levels. Therefore, this study would provide further insights into the relative value of information from corpora, learners, and teachers in the selection of the most useful words for EFL learners. It would indicate whether the following assumptions hold true in instructed contexts: (a) words are learned according to their frequencies, (b) words occurred frequently in corpora representing language use in different discourse types are perceived as being useful for EFL learners (corpora vs. teachers), and (c) words perceived as being useful by teachers are learned by learners. Importantly, unlike earlier studies, by examining all three factors in a single study, the present research would indicate the extent to which corpus-driven data and teachers influence the learning of high-frequency words in EFL contexts.

4. Research questions

1. Is there a relationship between lexical coverage and EFL learner knowledge of high-frequency words?

2. Is there a relationship between lexical coverage and EFL teacher perceptions of the usefulness of high-frequency words?

3. Is there a relationship between EFL teacher perceptions of the usefulness of high-frequency words and EFL learner knowledge?

4. To what extent do lexical coverage and EFL teacher perceptions of word usefulness predict learner knowledge of high-frequency words?

5. Methodology

5.1. Participants

The teacher participants were 25 EFL teachers who were native speakers of English, and 53 who were non-native speakers. Out of the non-native speaker group, 26 were Vietnamese while 27 were from various different countries. The native speakers of English had experience teaching English in countries in which English is used as a foreign language (e.g., China, France, Japan, South Korea). The Vietnamese EFL teachers had experienced teaching English to Vietnamese EFL learners in Vietnam. The EFL teachers from various different countries (e.g., Indonesian EFL teachers) had experience teaching English as a foreign language to learners who shared the same L1 background as them (e.g., Indonesian EFL learners) in their home countries (e.g., Indonesia). Sampling teachers from three groups would help to see whether teachers' familiarity with learners' language learning experience, contexts, and L1 is likely to influence vocabulary learning. All 78 teachers had experience teaching EFL learners from beginners to advanced

levels. The diversity in these teachers' L1 backgrounds, teaching contexts, and experience should result in a thorough picture of the perceptions of English language teachers about word usefulness for L2 learners. Further information about these teachers is presented in Appendix A.

The learner participants were 275 Vietnamese EFL undergraduate students from six universities in Vietnam. These participants shared features of learners studying foreign languages in many other contexts. They studied English in their home country where English is not the first or official language; their main contact with English was in language class at school but the exposure to English during the class periods may be limited in time, source, quantity, and quality (Muñoz 2008; Webb & Nation, 2017). Using learners having the same L1 and learning context as one group of teacher participants (Vietnamese EFL teachers) made it possible to examine the change in teacher perceptions of word usefulness according to their teaching context. It also allowed the researchers to conduct follow-up participant checking about their answers to increase the reliability of the results. The academic majors of the learners varied (Appendix B). They had studied English for 2-15 years, with an average of 8.92 (*SD*=2.18). The learners were divided into four groups (beginner, pre-intermediate, intermediate, advanced) according to their scores on Schmitt *et al*.'s (2001) Vocabulary Levels Test (VLT), which was delivered as part of their program entry tests (see Appendix C).

5.2. Target-words

973 headwords were randomly selected from Nation's (2012) BNC/COCA list as the target words. The BNC/COCA list was developed based on the frequency in the British National Corpus (BNC) and the Corpus of Contemporary American English (COCA), the two largest and best-known corpora of English. This list has been validated and widely used in vocabulary research (Dang & Webb, 2016; Dang *et al*., 2022; Nation, 2016).

5.3. Instruments

Ten surveys using a five-point Likert scale were created to investigate the teachers' perceptions

of the usefulness of the target-words for their students to perform basic functions in English.

Point 1 on the scale was labelled as the least useful and Point 5 the most useful. The instructions

were in English and put at the top of the surveys (Appendix D). The surveys were in an Excel

format and were emailed to each teacher (see Appendix E for the distribution of items in each

survey). Emailing surveys to participants allows researchers to collect data from a geographically

disparate population of teachers while causing minimal intrusion into their busy working lives

(Dörnyei & Taguchi, 2010). Moreover, if all 973 target-words had been included in an online

survey, it may have either discouraged the teachers from participating in this study or led to

fatigue effects. Distributing the target-words into 10 short surveys for teachers to do when they

had time solved this problem. Also, emailing the surveys to each teacher helped the researchers

to better manage the progress of each participant.

Fifteen Yes/ No tests were designed to measure learners' receptive knowledge of the form-

and-meaning relationship of the target-words. Form-and-meaning relationship was chosen

because it is the most important aspect of knowing a word and provides the foundation for

further development of other aspects of vocabulary knowledge (Nation, 2013). The Yes/ No

format was chosen because it is an appropriate format to measure a large number of target-words

with a large number of participants (Schmitt *et al*., 2011). It is not difficult to construct,

administer, or score, and allows a higher sampling rate for reliable estimation because a large

number of items can be tested in a limited period of time (Read, 2000). One limitation of the

Yes/ No format is face validity. The format does not require learners to demonstrate their

vocabulary knowledge, which may lead to the risk of test-takers not taking the test seriously

(Nation & Webb, 2011). Despite this criticism, a large number of studies have found strong positive correlations between tests using the Yes/ No test format with those using other formats such as multiple choice (e.g., Meara & Buxton, 1987), matching (e.g., Mochida & Harrington, 2006), and cloze (e.g., Fairclough, 2011). This suggests that Yes/ No format is a suitable format to measure test takers' receptive vocabulary knowledge.

To minimize learners' overestimation of their vocabulary knowledge, 480 pseudowords were included in the Yes/No tests in the present study. Pseudowords (e.g., *purfume*, *freath*) are items that look like real words in the language being tested (Meara & Buxton, 1987). The inclusion of pseudowords is based on the assumption that a person knowing all the words will tick 'Yes' to all the real words, but 'No' to all the pseudowords (Meara & Buxton, 1987). If the test-takers indicate that they know the pseudowords, their results will be adjusted accordingly. To deal with pseudowords, this study followed Schmitt *et al.*'s (2011) approach by excluding participants who checked more than 10% of pseudowords. The instructions were written in the learners' L1 and put at the top of the tests (Appendix F). See Appendix G for the distribution of items in each test. The tests had a paper-and-pencil format because it allows the first author to supervise the data collection, which increased the chances that the participants completed the tests and took the tests seriously. The supplementary materials of the present study can be freely downloaded on the IRIS Database; iris-database.org.

5.4. Corpora

Eighteen corpora of English were used in this study to determine the lexical coverage of the target-words in English spoken and written discourse (Appendices H and I). The number of tokens in each corpus ranged from 320,496 to 87,602,389. These corpora represented 10 different varieties of English. There was a good balance between the number of spoken (9) and

12

written corpora (9). Given their large number and variety, these corpora could provide an accurate assessment of the lexical coverage of the target-words from the perspective of corpus linguistics.

5.5. Collecting and analyzing the teacher and learner data

The data collection with the teachers was conducted in two stages. In the first stage, an official invitation was sent to EFL teachers through different channels such as teacher networks or face-to-face meetings. The first author set up one-on-one meetings with teachers who were interested in participating in the study to provide them with necessary information about the study and step-by-step instructions of the way to complete the surveys. In the second stage, the surveys were emailed to each individual teacher in an Excel attachment. Each teacher completed all ten surveys. Because of their busy working schedules, the teachers were given the flexibility to choose how often and how many surveys would be sent to them each time. The teachers downloaded the surveys, completed them, and emailed them back when they finished. To minimize the impact of the variation in the way that teachers responded to the surveys on the results of the study, the teachers were asked to complete the surveys as soon as possible but not to try to finish them all at the same time. After that, the results were checked to ensure that the teachers completed all sections in the surveys.

The learners were recruited on a voluntary basis. They received a small gift and their individual test scores with some interpretation, including their estimated vocabulary size and tips on how to improve their vocabulary knowledge. Each learner participant completed all 15 Yes-No tests over a period of two months. The tests were administered to the students during breaks between their English lessons and supervised by their instructors and the first author. All instructions were given in the learners' L1 so that they were clear about the study purpose and

the way to complete the tests. The learners were also informed that the tests were used for research purposes only and the test results would not affect their academic results. They were given as much time as they needed to complete the tests. Data inputting was done immediately after each test was collected.

To ensure the Yes/ No tests provided accurate estimation of vocabulary knowledge, data from 30 learners who ticked more than 10% of the pseudowords were excluded. It means only the data from 245 learners ticking no more than 10% of the pseudowords were included for analysis. Ten percent was found to be an acceptable percentage of checked pseudowords in Schmitt *et al*.'s (2011). Following Schmitt *et al*.'s (2011) approach, a series of independent-sample t-tests were conducted to compare the scores in the VLT of learners who ticked no more than 5% of the pseudowords in the Yes/ No tests, with those who ticked no more than 10% of the pseudowords. The results showed no significant difference in the overall VLT score for the 5%-set ($M=30.37$, $SD=24.94$), and the 10%-set, $M=31.98$, $SD=24.79$, $t(457) = -.69$, $p=.49$ (two-tailed), $\eta^2 = .0011$. As normality was in doubt, these results were supplemented with Mann-Whitney U test to confirm the results ($p = .44$). Similar analysis with the scores of the learners at each level of the VLT revealed the same results. Together, these results suggested that choosing 5% or 10% did not make any difference in the VLT mean scores. Moreover, a comparison between the VLT scores of the 245 learners who ticked no more than 10% of the pseudowords with the total number of target-words they indicated were known in the Yes/ No tests revealed that there was a linear relationship between the two variables. In particular, there was a strong correlation between the learners' scores in the Yes/No test and those in the VLT ($r=.93$). Given the high validity of the VLT, this strong correlation suggested that the data of the 245 learners who ticked

no more than 10% of pseudowords in the Yes/ No tests were accurate indicators of their knowledge of the target-words.

5.6. Calculating the lexical coverage in the corpora

To identify the lexical coverage of the target-words in the 18 corpora, the frequency of each target-word in each corpus was determined by running each corpus through Heatley, Nation, and Coxhead's (2002) RANGE program with the target-word serving as the base wordlist. Then, the coverage provided by the target-word in each corpus was calculated by dividing its frequency by the number of running-words in the corpus and multiplying by 100. Next, the mean coverage provided by the target-word in the 18 corpora was determined by adding its coverage in each corpus together, and then dividing by the number of corpora. Mean coverage was used rather than combined frequencies because combined frequencies would bias the results towards the findings of the largest corpora.

6. Results

Bootstrap Pearson product moment correlations were conducted to examine possible relationships between the lexical coverage, learner vocabulary, and teacher perceptions of word usefulness of each of the 973 target-words (see Table 1). Bootstrapping provides us with robust results because it considers the current sample as the population, randomly draws new samples from it, and runs the statistical test repeatedly through these samples (Plonsky, 2015). It has been used by He and Godfroid (2018) to explore the relationship between frequency and teacher perceptions of the usefulness of academic vocabulary. The interpretation of correlation values between 0 and 1 followed Plonsky and Oswald's (2014) benchmarks for interpreting *r* effect

sizes in L2 research; that is, r values "close to .25 be considered small, .40 medium, and .60 large' (p.889).

TABLE 1.

Bootstrapped descriptive statistics of each target-word

Variables	Mean	BCa 95% CI of mean		SD
		Lower	Upper	
Mean coverage	.003	.002	.003	.003
Proportion of learners knowing the words				
Beginners	.28	.26	.29	.27
Pre-intermediate	.55	.53	.57	.28
Intermediate	.75	.74	.77	.22
Advanced	.91	.90	.92	.13
All four groups combined	.49	.48	.51	.22
Teacher perceptions of word usefulness (mean scores)				
English native speaking EFL teachers	3.19	3.14	3.23	.67
Vietnamese EFL teachers	3.67	3.64	3.71	.60
EFL teachers from varying countries	3.53	3.50	3.56	.49
All groups combined	3.47	3.43	3.50	.55

In answer to Research Question 1, irrespective of the learner groups, there were always significant but very small correlations between lexical coverage and learner vocabulary

knowledge of high-frequency words. As shown in Table 2, a very small but significant

correlation was found between lexical coverage and learner vocabulary knowledge of the 973

target-words, $r = .20$, $p < .001$. Similarly, when each group of learners was examined, lexical

coverage and learner vocabulary knowledge were significantly and slightly correlated at $p < .001$,

but the effect sizes of the correlations were very small (r ranging from .17 to .22). This

significant correlation means that the higher lexical coverage of a word, the more likely it is

known by learners, and vice versa. However, the very small correlation between the two

variables means that this trend is only true to a degree.

TABLE 2.

Bootstrapped Pearson's correlations between lexical coverage and learner vocabulary knowledge

					BCa 95% CI	
Groups of learners	N	r	p	R^2	Lower	Upper
Beginner	133	.17	<.001	.03	.08	.25
Pre-intermediate	32	.19	<.001	.04	.12	.26
Intermediate	40	.22	<.001	.05	.17	.28
Advanced	40	.19	<.001	.04	.15	.25
All four groups combined	245	.20	<.001	.04	.12	.27

In answer to Research Question 2, Table 3 demonstrates that a very small but significant correlation was found between lexical coverage and teacher perceptions of word usefulness in the case of all teachers ($r = .23$, $p < .001$) and each group of teachers (r values ranging from .21 to .23, $p < .001$). The significant correlations mean that the higher coverage a word provides, the more likely it is perceived as being useful by teachers. The very small correlations indicate that it is only partially true that words with higher coverage are more likely to be perceived as being useful by teachers than words with lower coverage.

TABLE 3.

Bootstrapped Pearson's correlations between lexical coverage and the teacher perceptions of word usefulness

Groups of teachers	N	r	p	R^2	Bca 95% CI Lower	Upper
Vietnamese EFL teachers	26	.23	<.001	.05	.18	.29
English native speaking EFL teachers	25	.22	<.001	.05	.15	.29
EFL teachers from varying countries	27	.21	<.001	.04	.15	.28
All groups combined	78	.23	.001	.05	.17	.29

In answer to Research question 3, there is a close relationship between teacher perceptions of word usefulness and learner knowledge. Table 4 shows that the two variables largely correlated with each other, $r = .67$, $p < .001$. Vietnamese EFL teachers had the highest correlation with

18

learner vocabulary knowledge (*r*= .71), then the EFL teachers from different countries (*r* = .62), and the English native speaking EFL teachers (*r*= .56).

TABLE 4.

Bootstrapped Pearson's correlations between teacher perceptions of word usefulness and learners knowledge (All learners)

| | | | | | Bca 95% CI | |
Groups of teachers	*N*	*r*	*p*	R^2	Lower	Upper
Vietnamese EFL teachers	26	.71	<.001	.50	.67	.74
EFL teachers from varying countries	27	.62	<.001	.38	.58	.66
English native speaking EFL teachers	25	.56	<.001	.31	.51	.60
All groups combined	78	.67	<.001	.45	.62	.70

When the data of each learner group and each teacher group were compared, similar patterns were found. The *r* values ranged from .45 to .71, indicating medium to large correlations (see Appendices J to M). Of the three groups of teachers, irrespective of the groups of learners compared, Vietnamese EFL teachers always had the highest correlation with learner vocabulary knowledge (*r* from .58 to .71). Next came the EFL teachers from different countries (*r* from .52

to .65). English native speaking EFL teachers had the lowest correlation with learner vocabulary knowledge (*r* from .45 to .56).

The significant correlations between learner vocabulary knowledge and teacher perceptions means that the words perceived as being useful by teachers are more likely to be known by learners, and vice versa. The difference in the strength of the correlations between learner vocabulary knowledge and the perception of word usefulness of each teacher group suggests that the likelihood that learners learn the words indicated as being useful by Vietnamese EFL teachers is greater than that those indicated by EFL teachers from different countries, which is greater than those indicated by English native speaking EFL teachers.

In answer to the last research question, and with thanks to a suggestion from an anonymous reviewer, a mixed effects model was conducted using the lme4 package in R statistical platform to determine how well lexical coverage and teacher perceptions predicted learner knowledge of the target words, and which factor better predicted the vocabulary knowledge. The dependent variable was learner vocabulary knowledge. Lexical coverage and teacher perceptions were fixed effects. Teacher and item were random effects. In all cases, the variance information factor (VIF) scores of lexical coverage and teacher perceptions were around 1.0, suggesting no issues with multicollinearity. The whole model explained 33% of the variance (conditional $R^2 = 0.33$) while the fixed effects explained 29% of the variance (marginal $R^2 = 0.29$).

TABLE 5.

Mixed effects model analysis of the effect of lexical coverage and teacher perception of word usefulness on learner vocabulary knowledge

20

	b	SE	95% CI	df	t	p
(Intercept)	0.91	0.02	[0.88, 0.95]	1.99	55.05	0.0003***
Lexical coverage	0.01	0.002	[0.01, 0.01]	2915	4.78	1.86e-06 ***
Teacher perceptions of word usefulness	0.07	0.002	[0.07, 0.08]	2908	31.51	< 2e-16 ***

*p < .05, **p < .01, ***p < .001 (two-tailed)*

Table 5 showed that both lexical coverage and teacher perception significantly affected learner vocabulary knowledge. The latter had a greater effect than the former. As the teacher perceptions of word usefulness increased one unit, the learner vocabulary knowledge increased 0.07 units. As the lexical coverage increased one unit, the learner vocabulary knowledge increased 0.01 units. Similar results were found from the examination with each group of learners (see Appendices N to Q). Taken together, the findings indicate that both teacher perceptions of word usefulness and lexical coverage significantly affected vocabulary learning but not to the same extent. This means that learners are more likely to learn the words perceived as being useful by teachers than those occurring frequently in corpora.

7. Discussion

This study is the first to explicitly investigate the relationships between lexical coverage in corpora representing language use at large, EFL learner knowledge, and EFL teacher perceptions of the usefulness of high-frequency words. Therefore, it extends research on vocabulary and teacher cognition in several ways. First, it provides further insights into the current debate on the value of corpus-based high-frequency word lists for EFL learners. Second, it sheds light on the significant role of teachers' perceptions of usefulness on ultimate vocabulary learning in EFL

contexts, which is an underexplored area of both vocabulary research and teacher cognition research.

7.1. The value of corpus-based high-frequency word lists for EFL learners

It has been widely accepted among vocabulary researchers that corpus-based high-frequency wordlists are useful resources for foreign language learners (e.g., Nation, 2013; Schmitt, 2010). However, practitioners (e.g., Stein, 2017) point out that some items in these lists do not appear to be relevant to foreign language learners, and thus, raise the concern that teachers and learners may not see clearly the value of corpus-based wordlists to their teaching and learning. Drawing on data from a large number of corpora with a great degree of diversity, the vocabulary knowledge of EFL learners with different proficiency levels, and the perceptions of EFL teachers from various contexts, to some extent, this study provides empirical evidence to support the use of frequency-based criteria in corpora representing language use at large to select the most useful words for EFL learners. It found significant relationships between lexical coverage in corpora and (a) teacher perceptions of word usefulness and (b) learner vocabulary knowledge. This indicates that to some degree, the more frequently a word occurs in corpora representing language use at large, the more likely it is to be learned by EFL learners and perceived as being useful by EFL teachers. This finding can be explained by the nature of vocabulary acquisition and teacher cognition (i.e., what teachers think, believe, and know). Words with high lexical coverage tend to occur with higher frequency and wider range in language use at large (Ellis, 2002), and therefore, tend to be learned before those with lower lexical coverage. As for teacher cognition, teacher experiences as learners establish their cognition about learning and language learning, and this may continue to have an impact on teaching practice throughout their professional lives (Borg, 2006). Teachers' experience of learning and using the language as

22

native speakers (English native-speaking EFL teachers) or advanced EFL learners (the two groups of EFL teachers) may make the teacher participants aware of how often their learners may encounter a word in natural language use. This may affect their decisions about the usefulness of a word for their learners. Lexical coverage in corpora representing language use at large, learner knowledge, and teacher perception of word usefulness have been used to determine the most useful words for L2 learners. The significant relationships between lexical coverage and the other factors suggest that corpus-driven information (e.g., lexical coverage) is also consistent with the information from learners and teachers. Therefore, to some extent, the findings of this study highlight the importance of frequency-based criteria in the selection of the most useful words for EFL learners and the value of corpus-based high-frequency word lists for these learners.

However, this study also suggests that corpus-based high-frequency word lists may not have as large an impact on EFL vocabulary learning as might be expected. The very small correlations between lexical coverage and the other two factors indicate that it does not always hold true that the words occur frequently in corpora are to be perceived as being useful by EFL teachers and learned by EFL learners. The very weak relationship between lexical coverage and teacher perceptions of word usefulness is probably because factors other than lexical coverage also influence teachers' perceptions of word usefulness. Their experience of learning English as a foreign language (the two groups of EFL teachers) and teaching English for EFL learners (all three groups) may make the teacher participants aware that lexical coverage, and in turn frequency, is not the only factor that determines the degree of usefulness of a word for their students. Other situational factors (e.g., learning purposes, tests, curricula, materials) are also important (Borg, 2006). This assumption is supported by Dang and Webb's (2020) survey with

EFL teachers which revealed that textbooks were the most popular sources used by teachers to select words for their students (indicated by 87% of the teachers) while corpus-based wordlists were the least popular sources (used by 25% of the teachers). Meanwhile, research on vocabulary in EFL textbooks revealed that high-frequency words were poorly represented in these textbooks (e.g., Nguyen, 2020; O' Loughlin, 2012; Sun & Dang, 2020). The poor representation of high-frequency words in EFL textbooks and the influence of textbooks on teachers' word selection help to explain the weak correlations between lexical coverage and teacher perceptions of word usefulness.

The very weak relationship between lexical coverage and learner vocabulary knowledge may be due to the insufficient amount of input for these EFL students to learn high-frequency words and the influence of other factors rather than lexical coverage on vocabulary learning. To begin with, for lexical coverage and in turn frequency to impact vocabulary learning, a large amount of input is necessary for words to be encountered repeatedly (Webb & Nation, 2017). However, in many EFL contexts, learners have limited exposure to the target language (Muñoz, 2008). For many of them, the classroom may be the main environment to gain input and practice the target language (Laufer, 2001; 2003). Textbooks and English language teachers may be major sources of input (Horst, 2010; Richards, 2010). However, high-frequency words, which are indicated as being useful by corpora which representing language use on a large scale, are poorly represented in EFL textbooks (e.g., O'Loughlin, 2012; Sun & Dang, 2020) and the amount of incidental vocabulary learning through listening to teacher talk is relatively small (Jin & Webb, 2020). Meanwhile, this study found that learners were more likely to learn the words indicated as being useful by teachers than those indicated as being useful by the corpora, but the words indicated as being useful by the teachers did not always coincide with those indicated as being useful by

corpora. Another reason for the very small correlations between lexical coverage and learner vocabulary knowledge is that apart from lexical coverage, many interlexical factors (L1 influence and congruency) (Laufer, 1990) and intralexical factors (e.g., pronounceability, orthography, length, morphology, synformy) (Laufer, 1997) also influence vocabulary acquisition. The influence of these factors may reduce the potential for lexical coverage to strongly influence vocabulary learning in EFL contexts, resulting in relatively small correlations between lexical coverage and learner vocabulary knowledge.

The very small correlations between lexical coverage and the other factors may make readers question the pedagogical value of corpus-based high-frequency word lists for foreign language learners. It should be noted that although these correlations were very small, they were still statistically significant (all *p <.05*), and comparable to the correlation between frequency and teacher perception of academic formulas (*r=.26*, *p<.01*) found by Ellis *et al.* (2008). Moreover, it has been widely accepted by vocabulary researchers that it is useful for learners to learn items from corpus-based high-frequency word lists because learners only need to learn a small number of words. Knowledge of these words would enable learners to improve their comprehension and communication a great deal, which would further facilitate their future vocabulary development (e.g., Nation, 2013; Schmitt, 2010). What is unclear from earlier studies is the extent to which these words are known by learners and conform to teacher perceptions of word usefulness. Therefore, the very small correlations between lexical coverage and the other two factors found in this study should be considered potentially as an indication of the small impact of corpus-based high-frequency word lists on foreign language teaching rather than an indication that these lists have little value for foreign language learners. This finding provides some implications for researchers, teachers, and learners. To maximize the impact of corpus-based high-frequency

word lists on foreign language vocabulary learning and teaching, researchers should use information from teachers and learners to supplement frequency-based criteria in word list development and validation. In this way, corpus-based word lists could not only reflect the language features that foreign language learners may often encounter and use but also match the contextual and circumstantial realities of foreign language classrooms, thereby perhaps better serving the needs of learners and teachers. For foreign language teachers and learners, because knowledge of items from corpus-based high-frequency word lists allows foreign language learners to improve their comprehension and communication in various contexts (Nation, 2016), they need to be better aware of the value of these corpus-based high-frequency word lists.

In previous research, three factors have been used to determine the most useful words for L2 learners: lexical coverage in corpora representing language use at large, learner knowledge, and teacher perception of word usefulness. No studies have explicitly explored the relationships among these three factors. The present study helps to address this gap. It provides empirical evidence supporting the current trend in word list studies which uses learners' vocabulary knowledge (Dang *et al*., 2022; Brysbaert *et al*., 2020) and teacher perceptions of word usefulness (Dang *et al*., 2022; He & Godfroid, 2018) to supplement information from corpora. It also echoes the calls for more attention from teachers and learners to high-frequency words in foreign language contexts (e.g., Dang, 2020; Webb & Chang, 2012).

7.2. The role of teachers' perceptions of usefulness on ultimate vocabulary learning in EFL contexts

This study highlights the important role of teachers' perceptions of usefulness on vocabulary learning in EFL contexts. It found that teacher perceptions of the usefulness of high-frequency words was closely related to learner knowledge of these words and that teacher perceptions of

word usefulness better predicted learner vocabulary knowledge than lexical coverage in corpora representing language use at large. The close relationship between teacher perceptions of word usefulness and learner vocabulary knowledge can be explained by the nature of learning and teaching practice in EFL contexts and teacher cognition. Classrooms are the primary environment for many EFL learners to receive input and use the target language (Laufer, 2001; 2003). Therefore, teachers play an important part in vocabulary learning. The words that teachers introduce and the instruction that they give to learners may probably have a large impact on learners' vocabulary acquisition. What teachers do in classes is likely to be influenced by their cognition (Borg, 2006). If teachers think the words are useful for their learners, they are more likely to create learning opportunities to support their learners' acquisition of these words. It follows that the words that are perceived as being useful by teachers are more likely to be acquired by learners than those that are not.

This study found that teacher perceptions of word usefulness made a stronger unique contribution to explaining learner vocabulary knowledge than lexical coverage in corpora representing language use at large. This indicates that, in EFL contexts, teachers probably have a greater influence on what gets learned than natural exposure to the target language. This claim is supported by Laufer's (2001, 2003) studies which found that most of the words that these learners knew were more likely to come from deliberate learning than from incidental learning, and EFL learners gained more words through word-focused activities than through reading. This suggests that perhaps in many EFL contexts, learners may not have much exposure to the target language for incidental learning of the most frequent words to happen. As teachers play a key role in organizing learning activities, they may play a more important part in vocabulary learning in EFL contexts than natural exposure to the target language. Given that knowledge of items

from corpus-based high-frequency wordlists allows EFL learners to improve their comprehension and communication in various contexts (Nation, 2016) but EFL learners are more likely to learn the words considered as being useful by teachers than those having high frequencies in corpora, teachers should be aware of the great value of high-frequency words and provide many opportunities for students to get exposure to these words.

As the first study to examine the perceptions of teachers in a variety of contexts and considering it in relation to knowledge of learners from different proficiency levels and lexical coverage in various corpora, the present study provides useful evidence of the contribution of teachers to vocabulary learning in EFL contexts in particular and in foreign language in general. The vital role of teachers in high-frequency word learning found in the current study enriches our understanding of the complex process of learning vocabulary in a foreign language. It also highlights the value of research on teacher cognition of word usefulness, an underexplored area of vocabulary research. Such research can provide researchers and educational managers further insights into different aspects of vocabulary learning in foreign language contexts. Examining teacher perceptions of word usefulness like the present study can help to predict the words that are likely to be acquired by learners. Furthermore, as teacher perception of word usefulness is subject to many factors (Borg 2006), it can provide an implicit picture about the effect of different factors on vocabulary learning (e.g., culture issues, word frequency, learning materials). This information may then provide researchers and educational managers with a better idea of the constraints on vocabulary learning faced by a particular group of learners, and partially explain why a word is learned or not learned.

Another interesting finding in this study was the variation in correlations between learner knowledge and teachers from different contexts. Vietnamese EFL teachers always had the

highest correlations between perceptions of word usefulness and learner vocabulary knowledge; next came the EFL teachers from varying countries; the English native-speaking EFL teachers produced the lowest correlations. This might be because the Vietnamese teachers were the most familiar with the learner participants' language learning experience, contexts, and L1 while the English native-speaking EFL teachers were the least familiar. Similar to the learner participants, the Vietnamese EFL teachers acquired Vietnamese as an L1 when they were children and learned English as a foreign language when they were adults. Their knowledge of both Vietnamese and English may enable these teachers to predict the learning burden of an English word and best evaluate the value of the words for the learner participants.

The EFL teachers from varying countries ranked second in terms of familiarity with the language learning experience, contexts, and L1 of the learner participants. Like the learner participants, these teachers learned English as a foreign language. Additionally, these teachers experienced teaching in similar learning contexts as the learner participants. That is, their learners studied English in their home countries (e.g., Thailand), and were mainly taught by teachers sharing the same L1 with them. However, their unfamiliarity with the learner participants' L1 means that these teachers may not be able to evaluate the value of the words for these learners as well as the Vietnamese EFL teachers.

The English native-speaking EFL teachers were perhaps the least familiar with the learner participants' language learning experience, context, and L1. They learned English as their L1 when they were children. Although they may have learned other languages, these experiences may still be distant from the learner participants' L2 learning experience. Moreover, the learning contexts of the students of the English native-speaking EFL/ESL teachers were likely more varied. Some of their learners may have had similar learning contexts as the learner participants.

Yet, others may have had different learning contexts; that is, they studied English in English speaking countries and were mainly taught by teachers who did not share the same L1 with them. Such diverse contexts may affect the rating of the English native-speaking EFL teachers.

Taken together, this study highlights the importance for teachers to be familiar with learners' language learning experience (both L1 and L2) and contexts in vocabulary teaching. The findings of this study support the suggestion that teachers' knowledge of learners is fundamental in foreign language learning and teaching (Richards & Farrell, 2005; Richards, 2010). As little research has investigated this issue in vocabulary research and teacher cognition research (Borg, 2006), the present study provides a valuable contribution to the literature in both fields. Additionally, by showing that teachers who share the same L1 with learners are better at identifying the word value for learners than those who do not, this study reinforces the value of L1 in facilitating foreign language learning in general and vocabulary in particular (Schmitt, 2010).

This study took a quantitative approach by using surveys to explore teacher perceptions of word usefulness and check list tests to measure learner vocabulary knowledge. This approach allows researchers to objectively collect a large amount of data with a large number of participants, and therefore, optimizes the ability to generalize the findings (Bell & Waters, 2019). However, the quantitative approach cannot provide in-depth insight into the experience of each participant (Croker, 2009). For example, this study did not focus on why teachers thought a particular word useful for their students and why learners knew a particular word. Qualitative data from interviews or think-aloud would provide further insights into the reasons behind teachers' ratings of word usefulness and learners' vocabulary knowledge. This work was outside the scope of the current study. Moreover, this study only examined the vocabulary knowledge of

learners in one context (Vietnamese EFL context) and focused on English as a foreign language. It would be useful to replicate the study with learners in different contexts and other foreign languages and match the data from each group of teachers with a group of students from their same context.

8. Conclusion

Drawing on information from a large number of corpora, teachers, and learners with a great degree of diversity, this study provides empirical evidence for the use of corpus-driven information as the key criteria to identify the most useful words for EFL learners and the use of information from teachers and learners to supplement the that from corpora in the word selection. It also contributes to an underexplored area of vocabulary and teacher cognition research by (a) showing that teachers may have a greater influence on the learning of high-frequency words in EFL contexts than natural exposure to the target language and (b) highlighting the importance of teachers' familiarity with learners' language learning experience, contexts, and L1 in foreign language vocabulary learning.

References

Bell, J., & Waters, S. (2019). *Doing your research project: A guide for first-time researchers*. 7th edition. London: Open University Press, McGraw-Hill Education.

Borg, S. (2006). *Teacher cognition and language education*. London: Continuum.

Brezina, V., & Gablasova, D. (2015). Is there a core general vocabulary? Introducing the New General Service List. *Applied Linguistics*, *36*(1), 1–22. https://doi.org/10.1093/applin/amt018

Brysbaert, M., Keuleers, E., & Mandera, P. (2020). Which words do English non-native speakers know? New supernational levels based on yes/no decision. *Second Language Research*. https://doi.org/10.1177/0267658320934526

Coxhead, A. (2017). Academic vocabulary in teacher talk: Challenges and opportunities for pedagogy. *Oslo Studies in Language*, *9*(3), 29–44. https://doi.org/10.5617/osla.5845

Croker, R.A. (2009). An introduction to qualitative research. In J. Heighan & R.A. Croker (Eds.), Qualitative research in applied linguistics (pp.3-24). New York: Palgrave Macmillian.

Dang, T. N. Y. (2020). 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., & Webb, S. (2016). Evaluating lists of high-frequency words. *ITL-International Journal of Applied Linguistics*, *167*(2), 132–158. https://doi.org/10.1075/itl.167.2.02dan

Dang, T. N. Y., & Webb, S. (2020). Vocabulary and good language teachers. In C. Griffiths & Z. Tajeddin (Eds.), *Lessons from good language teachers* (pp. 203–218). Cambridge: Cambridge University Press.

Dang, T. N. Y., Webb, S., & Coxhead, A. (2022). Evaluating lists of high-frequency words: Teachers' and learners' perspectives. *Language Teaching Research*, *26*(4), 617–641. https://doi.org/10.1177/1362168820911189

Dang, Y. (2018). Teaching the most useful words of English. In J. I. Liontas, 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). Oxford: Wiley-Blackwell.

Dóczi, B., & Kormos, J. (2016). *Longitudinal developments in vocabulary knowledge and lexical organization* (pp. vii+-222). Oxford: Oxford University Press.

Dörnyei, Z., & Taguchi, T. (2010). *Questionnaires in second language research*. New York: Routledge.

Edwards, R., & Collins, L. (2011). Lexical frequency profiles and Zipf's law. *Language Learning*, *61*(1), 1-30. https://doi.org/10.1111/j.1467-9922.2010.00616.x

Ellis, N. C. (2002). Frequency effects in language processing. *Studies in Second Language Acquisition*, *24*(2), 143–188.doi:10.1017.S0272263102002024

Ellis, N., Simpson-Vlach, R., & Maynard, C. (2008). Formulaic language in native and second language speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL Quarterly*, *42*(3), 375–396.doi:10.1002/j.1545-7249.2008.tb00137.x

Fairclough, M. (2011). Testing the lexical recognition task with Spanish/English bilinguals in the United States. *Language Testing*, *28*(2), 273–297.doi:10.1177/0265532210393151

He, X., & Godfroid, A. (2018). Choosing words to teach: A novel method for vocabulary selection and its practical application. *TESOL Quarterly*.doi:10.1002/tesq.483.

Heatley, A., Nation, I. S. P., & Coxhead, A. (2002). *Range: A program for the analysis of vocabulary in texts*. http://www.vuw.ac.nz/lals/staff/paul-nation/ nation.aspx

Horst, M. (2010). How well does teacher talk support incidental vocabulary acquisition? *Reading in a Foreign Language*, *22*(1), 161–180.

Horst, M., & Collins, L. (2006). From faible to strong: How does their vocabulary grow?. *Canadian Modern Language Review*, *63*(1), 83-106.

Jin, Z., & Webb, S. (2020). Incidental vocabulary learning from listening to teacher talk. *Modern Language Journal*, *104*(3), 550–566. https://doi.org/10.1111/modl.12661

Laufer, B. (1990). Words you know: How they affect the words you learn. In J. Fisiak (Ed.), *Further insights into contrastive analysis* (pp. 573–593). Amsterdam: John Benjamins.

Laufer, B. (1997). What's in a word that makes it hard or easy: Some intralexical factors that affect the learning of words. In N. Schmitt & R. Carter (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 140–155). Cambridge: Cambridge University Press.

Laufer, B. (2001). Reading, word-focused activities and incidental vocabulary acquisition in a second language. *Prospect*, *16*(3), 44–54.

Laufer, B. (2003). Vocabulary acquisition in a second language: Do learners really acquire most vocabulary by reading? Some empirical evidence. *The Canadian Modern Language Review*, *59*(4), 567–587.doi: 10.3138/cmlr.59.4.567

Laufer, B., & Nation, I. S. P. (2012). Vocabulary. In S. M. Gass & A. Mackey (Eds.), *The Routledge handbook of secodn language acquisition* (pp. 163–176). London: Routledge.

Meara, P., & Buxton, B. (1987). An alternative to multiple choice vocabulary tests. *Language Testing*, *4*(2), 142–154.doi: 10.1177/026553228700400202

Milton, J. (2009). *Measuring second language vocabulary acquisition*. Bristol: Multilingual Matters.

Mochida, K., & Harrington, M. (2006). The Yes/No test as a measure of receptive vocabulary knowledge. *Language Testing*, *23*(1), 73–98.doi: 10.1191/0265532206lt321oa

Muñoz, C. (2008). Symmetries and asymmetries of age effects in naturalistic and instructed L2 learning. *Applied Linguistics*, *24*, 578–596.doi 10.1093/applin/amm056

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review*, *63*(1), 59–82.doi: 10.3138/cmlr.63.1.59

Nation, I. S. P. (2012). *The BNC/COCA word family lists*. Retrieved from http://www.victoria.ac.nz/lals/about/staff/paul-nation

Nation, I. S. P. (2013). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.

Nation, I. S. P. (2016). *Making and using word lists for language learning and testing*. Amsterdam: John Benjamins.

Nation, I. S. P., & Webb, S. (2011). *Researching and analyzing vocabulary*. Heinle, Cengage Learning.

Nguyen, C. D. (2020). Lexical features of reading passages in English textbooks for Vietnamese high-school students: Do they foster both content and vocabulary knowledge? *RELC*. https://doi.org/10.1177/0033688219895045

O'Loughlin, R. (2012). Tuning into vocabulary frequency in coursebooks. *RELC Journal*, *43*(2), 255 –269.doi: 10.1177/0033688212450640

Plonsky, L. (2015). *Advancing quantitative methods in second language research*. London:

Routledge.

Plonsky, L. & Oswald. F. L., (2014). How big is "big"? Interpreting effect sizes in L2 research.

Language Learning 64(4), 878–912.doi: 10.1111/lang.12079

Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.

Richards, J. C. (2010). Competence and performance in language teaching. *RELC*, *41*(2), 101–

122.doi: 10.1177/0033688210372953

Richards, J. C., & Farrell, T. S. C. (2005). *Professional development for language teachers:

Strategies for teacher learning*. Cambridge: Cambridge University Press.

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. New York:

Palgrave Macmillan.

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.doi: 10.1111/j.1540-

4781.2011.01146.x

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.doi:

10.1177/026553220101800103

Stein, G. (2017). Some thoughts on the issue of core vocabularies: A response to Vaclav Brezina

and Dana Gablasova: 'Is there a core general vocabulary?' Introducing the New General Service

List. *Applied Linguistics*, *38* (5), 759–763.doi: 10.1093/applin/amw027

Sun, Y., & Dang, T. N. Y. (2020). Vocabulary in high-school EFL textbooks: Texts and learner knowledge. *System*, *93*. https://doi.org/10.1016/j.system.2020.102279

Webb, S. A., & Chang, A. C.-S. (2012). Second language vocabulary growth. *RELC Journal*, *43*(1), 113–126. https://doi.org/10.1177/0033688212439367

Webb, S., & Nation, I. S. P. (2017). *How vocabulary is learned*. Oxford University Press.