



Deposited via The University of Leeds.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/238386/>

Version: Accepted Version

---

**Article:**

Dang, Y. (Accepted: 2026) Second language knowledge of academic collocations and its relationship with knowledge of academic words and general high-frequency words. *Studies in Second Language Acquisition*. ISSN: 0272-2631 (In Press)

---

**Reuse**

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

**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](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.

# **Second language knowledge of academic collocations and its relationship with knowledge of academic words and general high-frequency words**

## **Abstract**

Knowledge of academic collocations is important for academic communication. Examining its relationship with knowledge of academic words and general high-frequency words should provide insights into the relative value of prior knowledge of different kinds of single words for the development of collocational knowledge in academic contexts, thus expanding our understanding of vocabulary knowledge beyond the current conception of individual word knowledge. Yet no studies have addressed this issue, mainly due to the lack of validated tests for academic words and collocations. This study employed three pairs of newly developed tests to measure Vietnamese EFL learners' knowledge of academic collocations, academic words, and general high-frequency words. Results showed that at the form-recall level, both academic word and general high-frequency word knowledge were positively associated with academic collocation knowledge, with academic words having a stronger relationship. At the form-recognition level, only academic words and academic collocations were positively correlated.

*Keywords:* academic word; academic collocation; general word; vocabulary knowledge; vocabulary test; vocabulary knowledge framework

## **INTRODUCTION**

Nation's (2001, 2022) framework is probably the best-known model of vocabulary knowledge to date. However, scholars (e.g., Schmitt, 2023) have pointed out that this framework tends to focus primarily on single words. In fact, in Nation's framework, collocation is only considered as one aspect of single word knowledge. While knowledge of

single words may contribute to learners' ability to acquire collocations, treating collocations as part of single word knowledge may overlook the unique characteristics of multiword sequences. Earlier studies (e.g., Crossley et al., 2015) have shown that knowledge of multiword sequences is a good indicator of high language proficiency levels. Moreover, there is a growing body of evidence that multiword units are stored as a single unit in the mind (e.g., Ellis et al., 2008) and that the meaning of a multiword unit is not always transparent from the meaning of its single-word components (e.g., Martinez & Murphy, 2011). Not only does it not recognise collocations as a single unit, but Nation's framework also tends to neglect their relationships with other kinds of vocabulary. Although the association aspect in this framework recognises the inter-relations across lexical units, it focuses on relationships among single words rather than between multiword sequences and single words. With these considerations in mind, the primary focus of Nation's framework on single words leads to the need for exploring knowledge of multiword sequences and its relationships with knowledge of different kinds of single words. Nation and Webb (2011) have proposed a separate vocabulary knowledge framework for multiword sequences, which highlights the relationship between knowledge of multiword sequences and single words. However, only a small number of studies (Nguyen, 2023; Nguyen et al., 2024; Nguyen & Webb, 2017; Lu & Dang, 2023) have provided empirical evidence to support this framework. While these studies provide useful insights into the inter-relations between L2 knowledge of collocations and single general words by finding a significant relationship between the two, they have two limitations. First, they only measured vocabulary knowledge at the recognition level. It is unclear whether their finding still holds true in the case of recall. As recall is a higher level of vocabulary knowledge (Laufer & Goldstein, 2004), the finding might be different from that for recognition knowledge. Moreover, earlier studies either focused on the relationship between general collocations and general high-frequency words (i.e., the most frequent 3,000

words in general English) (Nguyen & Webb, 2017; Lu & Dang, 2023) or the relationship between academic collocations and general words (i.e., general high-frequency words together with words at lower frequency levels in general English) (Nguyen, 2023; Nguyen et al., 2024). It is unclear whether knowledge of academic collocations is also significantly related to knowledge of general high-frequency words and academic words or not, and if so, which relationship is stronger. It is worth investigating such relationships, because apart from academic collocations, knowledge of academic words (i.e., words occurring with high frequency across a wide range of subject areas in a large number of academic texts) and general high-frequency words is also important for academic communication (Nation, 2022). Examining the relationships between knowledge of academic collocations and each of these kinds of single words would provide further insights into the relative value of prior knowledge of single words for the development of collocational knowledge in academic contexts. This insight will expand our understanding of the nature of vocabulary knowledge beyond the current conception of individual word knowledge, as represented by Nation's framework. Yet no empirical studies have explored these relationships. This lack of research could be due to the limited number of validated tests to measure knowledge of academic words and collocations, especially at the recall level.

The present study will address this gap by using recently developed batteries of tests of academic collocations, academic words, and general words to measure L2 learners' knowledge of academic collocations and examine its relationships with knowledge of academic words and general words. By focusing on these relationships, the present study aims to provide empirical evidence to inform the expansion of the current conception of vocabulary knowledge to knowledge of multiword sequences, especially in academic contexts. Moreover, given that knowledge of academic collocations is important for effective academic communication but is difficult to acquire (Ackermann & Chen, 2013), findings

related to the relationships found between academic collocations and the two kinds of single words would provide teachers with insights into the relative value of academic words and general high-frequency words for the learning of academic collocations. This will inform their decision on which kinds of vocabulary that L2 learners should prioritise at different stages of their study to facilitate the development of academic collocation knowledge.

## **BACKGROUND**

### **Vocabulary knowledge frameworks of single words and multiword sequences**

One key concept in vocabulary research is what is involved in knowing a word. To date, Nation's (2001, 2022) framework is the best known. In this framework, word knowledge is classified into form (spoken, written, and word parts), meaning (form and meaning, concept and referents, associations), and use (grammatical functions, collocations, constraints on use). Each aspect is further divided into *receptive* and *productive* knowledge. Receptive knowledge commonly denotes vocabulary that learners can recognize and understand in reading and listening, while productive vocabulary refers to lexical items that they can use in their writing and speaking. However, as Read (2000) pointed out, in practice receptive knowledge has generally been measured with selected-response test formats (e.g., multiple-choice or word-definition matching), while constructed-response formats (e.g., gap-filling or cloze-type tasks) have been used to assess productive knowledge. Read (2000) argued that operationalising the two types of knowledge in this way was better described as a distinction between *recognition* and *recall* of vocabulary. Subsequent authors (Laufer & Goldstein, 2004; Schmitt, 2010) expanded it into a four-way division – form and meaning recognition, and form and meaning recall – which is now widely accepted. Recall is a more demanding measure of word knowledge than recognition and may be a better indicator of the learners' ability to retrieve the words for speaking and writing. However, there is also currently debate

whether a meaning recall format like translation is a more valid measure of vocabulary knowledge than a meaning recognition task such as multiple choice or matching (Stoeckel et al., 2021; Webb, 2021). For the purposes of the present study, we considered it important to measure both recall and recognition knowledge of the target vocabulary.

Nation's framework has been widely used in vocabulary research to inform vocabulary construct. However, several areas need further attention. First, although collocation features in this framework, it is only considered as one aspect of single word knowledge: *what word or types of words occur with this one?* or *what words or types of words must we use with this one?* Second, Nation's framework seems to focus primarily on intralexical features (i.e., features internal to a lexical unit itself such as spoken, written form, and word parts). Although interlexical features (i.e., the relations between one lexical unit and the others) receive some attention in this framework via the *association* aspect (i.e., *what other words does this make us think of? what other words could we use instead of this one?*), this aspect tends to focus on relations across single words rather than between multiword sequences and single words. Considering the importance of knowledge of multiword items for L2 proficiency (Schmitt, 2023), it would be useful to expand Nation's framework to include knowledge of multiword sequences and examine its relations with knowledge of different kinds of single words.

Recognizing this need, Nation and Webb (2011) developed a working framework of knowledge of multiword sequences in the same way as Nation's (2001, 2022) framework of single word knowledge to guide the design of multiword sequence research. This framework considers multiword sequences as a single unit whose knowledge involves the same aspects as that of single words. Moreover, it recognizes the value of the relationships between knowledge of collocations and knowledge of single words to some extent via the *association* aspect (*what other words... does this make us think of*). Compared to Nation's one (2001,

2022), Nation and Webb's (2011) framework is less known. Therefore, empirical evidence supporting this framework, especially the association aspect, is limited (see the section *Relationships between knowledge of collocations and single words* below).

## **L2 learners' knowledge of academic collocations**

Academic collocations can be defined as pairs of words (typically adjective-noun, verb-noun, and adverb-adjective) that occur with high frequency across a wide range of subject areas in a large corpus of academic texts (Ackermann & Chen, 2013). Although several studies have measured EFL learners' knowledge of general collocations (e.g., González-Fernández & Schmitt, 2020; Lu & Dang, 2023; Nguyen & Webb, 2017; Sonbul et al., 2023), to the best of our knowledge, only Nguyen (2023), Nguyen et al. (2024) and Lee (2024) have measured learners' knowledge of academic collocations.

As part of the validation of their Academic Collocation Tests (ACTs), Nguyen and colleagues (Nguyen, 2023; Nguyen et al., 2024) used these newly developed tests to measure academic collocational knowledge of 233 English-major EFL students in Vietnam and 110 ESL students in New Zealand at the form recognition and form recall levels. The test items are 60 collocations sampled from Ackerman and Chen's (2013) Academic Collocation List. When the data of the Vietnamese EFL learners were analysed together with those from the ESL learners, their participants scored 43.38% on the ACT Recall and 74.28% on the ACT Recognition. When analysed separately, the Vietnamese EFL learners scored 31.58% and 64.7% on these tests, respectively. These findings suggest that EFL learners' form recall knowledge of academic collocations is likely to be smaller than their form recognition knowledge. However, there was a flaw in the design of Nguyen et al.'s ACT Recognition test, which used the multiple-choice format, in that one word was repeated in all four options. Consequently, the test appears to measure knowledge of single words rather than

collocations. A study based on revised version of this test would provide better insight into EFL learners' knowledge of academic collocations.

Lee (2024) conducted a study with 205 ESL students from a university in the US. Similar to Nguyen et al. (2024), Lee developed a test battery to measure the participants' knowledge of 64 academic collocations. She found that on average the participants scored 28%-47% at the form recall level and 49%-63% at the form recognition level. However, Lee did not describe in detail how the target academic collocations were sampled for her tests nor reported how her test had been validated.

Taken together, Nguyen (2023), Nguyen et al. (2024), and Lee (2024) have indicated L2 learners' knowledge of academic collocations varies depending on educational contexts, and their recall knowledge tends to be smaller than their recognition knowledge. However, the tests used in these studies had limitations in their design and validation procedures, which made it desirable to validate the collocation tests to be used in the present study with learners from the target population.

### **Relationships between knowledge of collocations and single words**

Investigating the relationship between prior knowledge of single words and academic collocation knowledge is important because earlier studies have found that prior knowledge of single words positively predicts incidental learning of collocations from reading (e.g., Vilkaite, 2017), reading while listening (e.g., Vu & Peters, 2021), and viewing (e.g., Puimège & Peters, 2020). Therefore, several studies have examined the relationships between L2 learners' knowledge of collocations and single words. Nguyen and Webb (2017) conducted a study with Vietnamese EFL English-major students. They used Webb et al. (2017) Updated Vocabulary Levels Test to measure the participants' knowledge of the connection between form and meaning of general high-frequency words at the recognition level and created a

Receptive Collocation Test to measure their knowledge of the collocations of general high-frequency words, also at the form recognition level. They found a significant relationship between the participants' knowledge of single words and collocations ( $r = 0.67$  in the case of verb-noun collocations and  $r = 0.70$  in the case of adjective-noun collocations). Later Lu and Dang (2023) replicated Nguyen and Webb's study with postgraduate students in China and also found a significant correlation between their participants' knowledge of the two kinds of vocabulary.

While Nguyen and Webb (2017) and Lu and Dang (2023) focused on collocations of general high-frequency words, Nguyen and colleagues (Nguyen, 2023; Nguyen et al., 2024) examined academic collocations. They used Nation and Beglar's (2007) Vocabulary Size Test (VST) to measure learners' meaning recognition knowledge of general words and their newly developed Academic Collocation Tests (ACTs) to measure these participants' knowledge of academic collocations at both form recall and form recognition levels. Data were collected from English-major EFL students in Vietnam and ESL students in New Zealand. When the data from the two groups of learners were analysed together, the participants' VST scores were moderately correlated with both their ACT recognition scores ( $r = 0.53$ ) and their ACT recall scores ( $r = 0.52$ ). When the data of the EFL learners were analysed separately, the correlations with the VST test were somewhat lower: ACT Recall  $r = 0.50$  and ACT Recognition  $r = 0.41$ . These findings indicated a significant relationship between knowledge of general words and academic collocations. However, Nguyen and colleagues measured knowledge of general words only at the recognition level. As recall is a higher level of vocabulary knowledge (Aviad-Levitzky et al., 2019; Laufer et al., 2004), looking at both the recognition and recall levels would provide better insights into the relationships between knowledge of the two kinds of vocabulary. Moreover, Nguyen and colleagues did not assess knowledge of academic words although this knowledge is also

important for academic communication (Nation, 2022). Examining their relationship with knowledge of academic collocations would provide useful insights into the relative value of knowledge of general high-frequency words and academic words in the acquisition of academic collocations. Yet no studies have investigated these particular relationships.

### **Research aims**

The present study aims to address the gaps identified in the literature review by investigating learners' knowledge of academic collocations at the form recall and form recognition levels and then exploring how this knowledge is related to their knowledge of both general high-frequency words and academic words.

The present study focused on form recall and form recognition because in practical terms, it reflected what was measured by the item types in the tests available to us, especially the Academic Collocation Tests (please see details in the *Original tests* section below).

Moreover, in the spectrum of knowledge of form-meaning links, form recall is the most demanding while recognition knowledge is the least demanding, with no significant difference in the degree of difficulty between form recognition and meaning recognition (Aviad-Levitzky et al., 2019; Laufer et al. 2004).

### **METHODOLOGY: PRELIMINARY STUDY**

Before undertaking the main study, we conducted an extensive preliminary study to evaluate the quality of the tests that we proposed to employ and to judge their suitability for use with the target population for research purposes: undergraduate students in Vietnam. The motivation for this initial work was to respond to recent calls from several scholars (Read, 2023; Schmitt et al., 2020) to apply to vocabulary research the professional standards for test development and validation that are accepted practice in the field of language assessment.

This includes validating tests for a defined purpose, in terms of how the results will be used, and tailoring the tests for a specified population of learners.

The preliminary study was carried out in three stages, each of which resulted in revisions of the test material: an expert review of the original tests; tryout of the revised tests with a small group of learners; and a more extensive trial with a larger, more representative group.

### **The original tests**

We set out to measure knowledge of academic collocations, academic words, and general high-frequency words at both recall and recognition levels. There are few published vocabulary tests that meet these requirements, particularly ones having parallel recall and recognition versions. However, we were able to identify three paired tests that could be revised and adapted as necessary for use in our project: Academic Collocation Tests (Nguyen et al., 2024), Academic Vocabulary Tests (Pecorari et al., 2019, 2025), and New Computer Adaptive Tests of Size and Strength (CATSS) (Aviad-Levitzky et al., 2019).

The Academic Collocation Tests (ACT) (Nguyen et al., 2024) were developed to measure knowledge of academic collocations at the form recall and form recognition levels (see Figure 1 for examples of the ACT items). There were 60 items in each test. These items were sampled from Ackerman and Chen's (2013) Academic Collocation List. Nguyen et al. followed the now standard procedures for vocabulary test development and then conducted a validation study following Kane's (2013) argument-based approach with English majors in Vietnam and ESL university students in New Zealand.

[INSERT FIGURE 1 HERE]

*Recall test:*

The two approaches have a **si** \_\_\_\_\_ **di** \_\_\_\_\_. One uses interviews while the other does not. (large variation)

*Recognition test:*

1. The two approaches have a \_\_\_\_\_ \_\_\_\_\_. One uses interviews while the other does not.
  - A. significant difference
  - B. significant effect
  - C. significant finding
  - D. significant number

**FIGURE 1.** Example of the Academic Collocation Test items (Nguyen et al., 2023)

The Academic Vocabulary Tests (AVT) (Pecorari et al., 2019, 2025) were designed to measure knowledge of academic words at the form recall and form recognition levels (see Figure 2 for examples of the AVT items). They were based on the Academic Vocabulary List (AVL) compiled by Gardner and Davies (2014) from the Corpus of Contemporary American English (COCA). The recall AVT consists of 52 target items while the recognition test has 57 items. These tests have been validated with EFL undergraduate students at Swedish universities.

[INSERT FIGURE 2 HERE]

*Recall test:*

1. Throughout his career he showed the qualities of dedication and comm_____.
---

*Recognition test:*

- |                                    |               |
|------------------------------------|---------------|
| ___ a promise to do something      | a. colleague  |
| ___ a thing that somebody has made | b. commitment |
| ___ someone who works for you      | c. creation   |
|                                    | d. experiment |
|                                    | e. flow       |
|                                    | f. percentage |

**FIGURE 2.** Example of the Academic Vocabulary Test items (Pecorari et al., 2019; 2025)

The New Computer Adaptive Tests of Size and Strength (CATSS) (Aviad-Levitzky et al., 2019) were designed to measure knowledge of general words. These parallel tests measure learners' vocabulary knowledge of form and meaning at both recall and recognition levels. Each test consists of fourteen 10-item components corresponding to 14 word-frequency levels of general English vocabulary, as defined by a combination of frequency lists derived from the British National Corpus (BNC) and the Corpus of Contemporary American English (COCA). These tests have been validated by Aviad-Levitzky et al. (2019) with students at a university in Israel following the argument-based approach. As the name indicates, the CATSS battery is primarily designed as a computer-adaptive instrument, but each parallel test can also be administered as separate, paper-based tests (Aviad-Levitsky et al. 2019), which was more appropriate for our research context. For our purposes, we selected only the

form recall and form recognition tests of CATSS (see Figure 3 for examples of the CATSS items).

[INSERT FIGURE 3 HERE]

*Recall test:*

I always w\_\_\_\_\_ to work. (go on foot).

*Recognition test:*

(1) I always \_\_\_\_\_ to work. (go on foot)  
a. move            b. walk            c. drive            d. jump

**FIGURE 3.** Example of items in the CATSS form recall and form recognition (Aviad-Levitzky et al., 2019)

### **Expert review**

The first stage of the validation was a review of the tests by advanced users of academic English: an L1 user of English who is a professor in Applied Linguistics at a university in an English-speaking country, and four L1 users of Vietnamese who had either completed or were currently doing PhD studies in Applied Linguistics in English-speaking countries.

Given their academic and professional experiences, these five experts were assumed to have a solid knowledge of the academic collocations, academic words, and general words needed to function effectively in academic and general communication. Moreover, the four Vietnamese experts had experience teaching English to Vietnamese university students from similar backgrounds as the target participants and thus had a good understanding of the vocabulary

levels of Vietnamese EFL students, so that they could comment on the relevance of the test items for the target participants. Results of this review indicated that, while the other tests seemed to be appropriate, ACT Recognition, AVT Recall, CATSS Recognition and Recall needed to be revised further to function effectively for the purposes of this study.

The test that required the most extensive revision was ACT Recognition. As shown in the sample item in Figure 1 above, one of the two words in the collocation was repeated in all four options of these multiple-choice items. This meant that in effect the items were testing only one of the two words in the collocation. When the four Vietnamese experts worked through the items, they reported that in many cases they looked only at the different words in the four options and selected the option that made sense in that context.

To revise ACT Recognition, the second author drafted revised test items following these five guidelines<sup>1</sup>:

1. The head word of each option (both correct answer and distractors) should match the context in the stem sentence. Otherwise test takers might rely on the meaning of the head word to guess the answer or eliminate distractors.
2. General collocations could be used as distractors but only the target (academic) collocation could fit in the stem sentence.
3. No words were repeated in the four options. One test-taking strategy is to look for a repeated word in a set of multiple-choice options.
4. Meanings of the phrases used as distractors should be similar to that of the target collocation. Otherwise, test takers could eliminate a distractor based on the meaning of its components.

5. The frequency of words in the distractors should be the same as the frequency of words in the target collocations to reduce the possibility of blind guessing. The frequency was checked against Nation's (2016) BNC/COCA 25,000 word lists.

When it was difficult to find suitable distractors satisfying all of these requirements, the guidelines were applied with flexibility. Academic collocations were checked in Ackerman and Chen's (2013) Academic Collocation List and the academic section of COCA. Once the second author had completed the first draft of the test, it was commented on by the other authors. The second author revised the test items based on the feedback.

Regarding AVT Recall, the main revision was to write new items for five target words – *allot*, *coherence*, *formulate*, *requisite*, and *subset* – which had been included in the AVT Recognition, but the original authors had found it difficult to compose suitable stem sentences for these words in their working version of the AVT recall (Percorari, personal communication, 20 January 2022). Since it was important for the design of the present study to test all the target words at both the recognition and recall levels, the first author wrote items for the five words by referring to example sentences in the Oxford English Dictionary and concordance lines in the academic section of the COCA corpus. The draft items were reviewed by the other four authors and revised based on their comments. A second significant revision was to add definition clues to all the AVT Recall items for consistency with the other recall tests (see the sample items in Figures 1-3 above). Other minor revisions in the interests of clarity involved revising the stem sentences of five items and adding more letters to the gaps in three other items.

Regarding CATSS, the second author checked all items in the CATSS recognition test and suggested shortening and simplifying several items and made changes in the CATSS recall test accordingly. This decision was in line with Carr's (2011) suggestion that stem sentences

should not contain too much information but should be restricted to the information needed to pose the question.

### **Tryout**

Once all the authors had approved the revisions, the six tests were tried out with 24 students from a similar cohort to the target participants. These students were divided into six groups (four students per group). Each group completed one test version. To examine the performance of learners at different language proficiency levels, each group consisted of both high and lower-level students, based on the evaluation of their teachers. Once the test scores had been collected, they were analysed both quantitatively and qualitatively. Quantitatively, we compared the mean scores of the recall and recognition groups for each test and examined the scores of individual students. Results showed that although the proficiency levels of participants in the two groups were comparable, the recall group always had lower scores than the recognition group, indicating that the participants' scores on the tests followed the typical lexical profile: L2 learners' recall knowledge lags behind their recognition knowledge. Moreover, participants rated as having higher language proficiency by their teachers mostly had higher scores on the tests than those with lower proficiency. Additionally, the participants' scores on the CATSS decreased from the 1<sup>st</sup> to 14<sup>th</sup> 1,000-word levels. This pattern suggested that overall, the CATSS scores reflected the typical lexical profile. Together the quantitative analyses of the overall test scores indicate that the revised CATSS, AVT, and ACT tests were suitable for the target participants.

We also examined the scores of each test item closely. While the results of this qualitative analysis largely supported the findings of the quantitative analyses, they also revealed a number of test items that needed further consideration. To deal with these items, first, we referred to our L2 teaching and learning experiences to explain why the participants provided certain answers to these items. If we could not identify the reasons for the participants'

answers, we conducted follow-up checking with them. After that, we discussed among the team whether we should revise the items or not. Items that did not need further revision included those based on target words (a) that were likely known by the participants from their textbooks/learning materials, or (b) that are cognates in their L1. There were also cases where the participants reported selecting an incorrect answer by mistake. Revisions were made in the stem sentences, clues, or distractors, as appropriate.

### **Piloting**

The revised versions of the tests were then piloted with 200 students who also came from the target population. Half of these students completed the recall versions of the tests while the other half completed the recognition versions. They took the tests in class time. Then, students were invited to join follow-up interviews to investigate whether the test scores accurately represented the participants' knowledge of target words or perhaps reflected the effects of test taking strategies. Fifteen students responded to the invitations, but 12 were available for the interviews eventually. Semi-structured individual interviews lasting up to two hours were conducted with these students outside of class time. The interviews were conducted by the first author via her institution' approved apps (Microsoft Teams).

Recordings of the interviews were stored in Microsoft Stream, password-protected and accessible only by the first author. Initially we planned to ask the participants to complete all three tests. However, we realised that three tests were too much for a two-hour session, leading to fatigue effects. Therefore, after the participants had completed one test, they were asked if they were still willing to complete a second test and they all consented to do so.

Although each participant only completed two tests, we made every effort to ensure that each test was completed by multiple students. In this way, we could obtain evidence for the validity of the tests without asking for too much time from the participants. When completing

each test in the interview, the participants went through all items in the test and recalled how they had responded to each one in the in-class test sessions.

It was not appropriate to apply a full validity framework to this kind of study because the tests were being evaluated for research purposes rather than to make pedagogical or other decisions about the students. Following the lead of researchers such as Schmitt et al. (2020) and Nguyen et al. (2024), for the analysis we drew selectively on two of the inferences from Kane's (2013) argument-based framework: Evaluation and Generalization. These inferences were the most appropriate ones for the purposes of our study.

Evidence for the Evaluation inference was collected from the Rasch analysis<sup>2</sup> of the in-class test scores (item measures and item fit) and interviewees' opinions about the tests. Results of the Rasch analysis of items showed that the mean item difficulty values for the recall tests are around 1.0 and 2.0 (see Table 1). However, the Min and Max values indicate that these tests include a range of items with different levels of difficulty. Regarding the recognition tests, the means of the item difficulty and person measures are close to 0 (Table 2). This indicates that these tests were generally appropriate to the participants' ability levels. Findings of the items measure analysis were supported by the Wright maps. Result of the Rasch analysis of item fit shows that the number of underfit items in each of the examined tests was fewer than five, which is within an acceptable limit. Our qualitative analysis of the in-class test and interview data related to each of these items revealed no obvious problems (see Appendix 1 in our Supplementary materials via <https://osf.io/3s8wx/overview> for examples of how we dealt with the underfit items).

[INSERT TABLES 1 & 2 HERE]

**TABLE 1.** Rasch Measures of the ACL, AVT, and CATSS Recall in Logits

	Scoring scheme	ACL		AVT		CATSS	
		<i>Mean (SD)</i>	<i>[Min, Max]</i>	<i>Mean (SD)</i>	<i>[Min, Max]</i>	<i>Mean (SD)</i>	<i>[Min, Max]</i>
Item	Strict	1.81 (1.74)	[-2.09, 5.22]	2.20 (1.98)	[-2.54, 5.51]	1.36 (2.59)	[-5.61, 5.32]
difficulty	Lenient	1.77 (1.81)	[-2.39, 5.33]	1.92 (2.04)	[-2.50, 5.45]	1.18 (2.69)	[-5.61, 5.36]
Person	Strict	0.003 (0.47)	[-4.40, 3.72]	-0.02 (0.49)	[-4.32, 3.47]	-0.02 (0.28)	[-3.83, 2.53]
measure	Lenient	0.006 (0.48)	[-4.61, 3.74]	-0.02 (0.45)	[-4.53, 3.32]	-0.02 (0.28)	[-3.88, 2.84]

**TABLE 2.** Rasch Measures of the ACT, AVT, and CATSS Recognition in Logits

	ACT		AVT		CATSS	
	<i>Mean (SD)</i>	<i>[Min, Max]</i>	<i>Mean (SD)</i>	<i>[Min, Max]</i>	<i>Mean (SD)</i>	<i>[Min, Max]</i>
Item difficulty	-0.09 (0.77)	[-1.91, 2.24]	-0.56 (1.40)	[-4.20, 3.54]	-0.58 (1.72)	[-4.98, 2.72]
Person measure	-0.03 (0.90)	[-2.11, 2.26]	-0.01 (0.35)	[-5.23, 2.84]	-0.01 (0.22)	[-2.37, 2.35]

Results of the follow-up interviews supported the findings of the Rasch analysis of the in-class test results. Overall, the participants were positive about the test and they indicated that test-taking time was sufficient. When asked to comment on each item of the tests, their responses showed that the instruction, stem sentences, initial letter hints, and clue definitions were clear for most items. They felt that the test reflected their vocabulary knowledge well. For example, one interviewee commented as follows:

*I think the tests are easy to understand. They are also valid. As most of the answers I guessed were incorrect, the tests are pretty accurate...*

In addition to the Evaluation inference, we also collected evidence for the Generalisation inference, which was based on the examination of test reliability. Table 3 presents the reliability estimates of the tests. Regardless of the test format and scoring scheme, the Cronbach’s alpha and Rasch reliability of the academic collocation recall was always around 0.90, indicating a very high level of reliability.

[INSERT TABLE 3 HERE]

**TABLE 3.** Reliability estimates of the tests

Test	Cronbach’s alpha	Rasch item reliability	Rasch person reliability
ACT Recall (strict)	0.99	0.99	0.92
ACT Recall (lenient)	0.99	0.99	0.94
ACT Recognition	0.90	0.99	0.89
AVT Recall (strict)	0.90	0.99	0.89
AVT Recall (lenient)	0.90	0.99	0.90
AVT Recognition	0.94	0.93	0.93
CATSS Recall (strict)	0.95	0.99	0.96
CATSS Recall (lenient)	0.95	0.99	0.96
CATSS Recognition	0.95	0.95	0.95

## **METHODOLOGY: MAIN STUDY**

### **Research questions**

Together the findings of the preliminary study indicated that the revised versions of the three paired tests were suitable for the target participants of our study, and they met the requirements of technical quality. Therefore, they were used as the instruments to measure knowledge of academic collocations, academic words, and general high-frequency words in the main study<sup>3</sup> to answer these research questions:

1. What is EFL learners' knowledge of academic collocations at the form recall as compared to the form recognition level?
2. To what extent is their knowledge of academic collocations associated with their knowledge of general high-frequency words and knowledge of academic words? If so, which relationship is stronger?

### **Participants**

The participants were undergraduate EFL students from a Foundation year program for non-English-major students and a four-year program for English-major students in Vietnam. The Foundation program lasts for one year before students spend the next three years on discipline-specific EMI courses as part of their joint degree in business between a Vietnam university and a UK university. The program is designed to help students to develop their general and academic language skills and literacies. The English-major students' program runs for four years over eight semesters. The first three semesters are devoted mainly to English language skills courses which focus on academic reading, listening, writing, and speaking. The remaining five semesters consist of disciplinary degree courses in English (translation, language teaching, culture, and literature). In addition to English-medium courses, this program also offers several Vietnamese-medium courses during the first three

semesters. Obviously, it is highly desirable for students in these programs to develop an adequate knowledge of English vocabulary for both receptive and productive use if they are to succeed in their subsequent studies. Students typically enter their program after having studied English in school as a compulsory subject from the age of 11. Upon entry their proficiency in English is generally around the A2 to B1 levels on the Common European Framework of Reference (CEFR) for the Foundation year students and the B1 level for the English-major students.

Originally 310 students participated in the study. Only data of 252 students who completed all the tests were used for the analysis. The original participants were randomly assigned to either the recall group or the recognition group (Table 4). The recall group completed the form recall version of tests of academic words, academic collocations, and general words while the recognition group completed the form recognition version of these tests. Each student only took one version of the tests to avoid testing effects. Results of the Mann Whitney U-test<sup>4</sup> showed that the two groups were comparable to each other in terms of the length of formal English instruction ( $p = 0.77$ ) and university entrance exam scores ( $p = 0.19$ ).

[INSERT TABLE 4 HERE]

**TABLE 4.** Participants in the recall and recognition groups

Group	Recall	Recognition
Foundation Year	27	25
Year 1	31	28
Year 2	21	30
Year 3	25	17
Year 4	30	18
Total	134	118

## **Instruments and procedure**

The instruments for the main study were the versions of the six original tests that had been revised and validated in the preliminary study (see the Appendix 2 in our Supplementary materials via <https://osf.io/3s8wx/overview> for these instruments). In response to an issue raised by a reviewer, we examined the overlaps between items in the test of academic collocations (Academic Collocation Test) with those in the tests of academic words and general high-frequency words (Academic Word Test and CATSS3000). Only one item in the Academic Collocation Test (*homogeneous group*) has a component as a test item in the Academic Word Test (*homogeneous*), and only two items (*initial phase*, *offer an opportunity*) in the Academic Collocation Test has a component as a test item in the tests of general high-frequency words (CATSS3000) (*initial*, *opportunity*). Therefore, we believe that overlaps between items across tests were unlikely to bias the findings of our study.

The data collection was conducted in class time in the middle of the participants' academic year. All the tests were administered as paper-based instruments. Students were given unlimited time to complete the tests, but it normally took from 25-50 minutes to complete each one (See Appendix 3 via <https://osf.io/3s8wx/overview> for the instruments and typical time to complete them).

The tests were administered and supervised by the third and fifth authors to ensure the students took the tests seriously. Instructions were provided in the participants' L1 (Vietnamese). Moreover, we stated as part of the instructions that the students should let us know if they do not understand any steps in how to complete the tests, as well as the meaning of any stem sentences, or clue definitions. Since no students raised any problems, we are confident that the test instruction, stem sentences, and clue definitions were clear to them.

To avoid fatigue effects, the data were collected over two face-to-face sessions. In the first session, the participants completed a short questionnaire to report their university entrance English Exam score and their length of formal English instruction and the test of general words, including general high-frequency words (CATSS). In the second session, they completed the test of academic words (AVT) and the test of academic collocations (ACT).

### **Scoring and analysis**

The raw data were input to Excel spreadsheets by the third and fifth authors and then marked by the first author. The recognition tests were marked dichotomously. Selected options which were the same as the answer key were coded as correct answers while those which were not the same or left blank were coded as incorrect answers. To obtain a fuller picture of the participants' recall knowledge of academic collocations, academic words, and general words, the marking of the recall tests was based on two scoring schemes (strict and lenient) and followed Read and Dang's (2022) procedure (see Appendix 4 via <https://osf.io/3s8wx/overview> for the detailed description of the scoring schemes). After the scoring had been completed, several linear mixed-effects models were made with the lmer package in R. In these models, university level and class were set as the random factors. Continuous variables (the academic collocation test score, the academic word test score, the general word test scores, the university entrance English Exam scores, and length of formal English instruction) were centred and scaled.

To examine the relationship between knowledge of academic words and general high-frequency words, we created three models:

- Model 1 for scores on the recall tests using the strict scoring scheme
- Model 2 for scores on the recall tests using the lenient scoring scheme
- Model 3 for scores on the recognition tests

In each model, the academic collocation test score was the dependent variable, whereas the academic word test score, the general high-frequency word score, university entrance English Exam, and the length of formal English instruction were fixed effects.

Before running these linear mixed-effects models, we also checked the correlations and multicollinearity of fixed effects. The results showed that all recall test scores had large correlations ( $r$  ranging from 0.60 to 0.73) while all recognition test scores had medium correlations ( $r$  ranging from 0.31 to 0.57) (see Appendix 5 via <https://osf.io/3s8wx/overview>).

The interpretation of the correlation size followed Plonsky and Oswald's (2014) guidelines.

Given the significant correlations of these test scores, we checked the Variance Inflation Factors (VIF) scores to see if multicollinearity was a problem. Following Öksüz et al. (2021), we applied the strict VIF cut-off score (i.e., VIF score above 5.0 is considered as potential multicollinearity problem). In our study, all VIF scores were between 1.11 and 2.95.

Therefore, multicollinearity was unlikely to be a problem. We then ran the linear mixed-effects models.  $R^2$  was calculated to see how well the models explain the variance in the dependent variable. The standardized coefficients ( $\beta$ ) were computed to interpret effect sizes of the fixed effects. The interpretation of the effect size followed Cohen's (2013) guideline: small ( $0.10 \leq \beta < 0.30$ ), medium ( $0.30 \leq \beta < 0.50$ ), and large ( $\beta \geq 0.50$ ).

## RESULTS

Table 5 presents the mean scores of the recall and recognition groups on the tests of academic collocations, academic words, and general high-frequency words, respectively. When the mean scores were converted to percentages, on average, the recall group scored 28.45% on the ACT, 21.04% on the AVL, and 72.71% on the CATSS 3000<sup>5</sup> if the strict scoring scale was employed. If the lenient scoring scheme was used, their mean scores on each test were slightly higher: 29.5% (ACT), 24.12% (AVL), and 75.30% (CATSS 3000). Compared to the

recall group, the recognition group had higher mean scores on the tests: 47.12% (ACT), 50.86% (AVT), and 87.97% (CATSS 3000).

[INSERT TABLE 5 HERE]

**TABLE 5.** Means (SD) of the recall and recognition group scores on the ACT, AVT, and CATSS 3000

Vocabulary test	Maxi mum score	Raw score			Percentage score (%)		
		Recall ( <i>N</i> = 134)		Recognition ( <i>N</i> = 118)	Recall ( <i>N</i> = 134)		Recognition ( <i>N</i> = 118)
		Strict	Lenient		Strict	Lenient	
ACT	60	17.07 (11.63)	17.70 (12.12)	28.27 (11.03)	28.45 (19.38)	29.50 (20.20)	47.12 (18.38)
AVT	57	11.99 (7.59)	13.75 (7.86)	28.99 (13.01)	21.04 (13.32)	24.12 (13.80)	50.86 (22.82)
CATSS 3000	30	21.81 (5.24)	22.59 (5.36)	26.39 (2.87)	72.71 (17.46)	75.30 (17.87)	87.97 (8.58)

Data of the recall group were used to examine the relationships between form recall knowledge of academic collocations and knowledge of each kind of single words (academic words and general-high-frequency words). Correspondingly, data of the recognition group were used to investigate the effects of form recognition knowledge of academic collocations with knowledge of each kind of single words.

Table 6 presents the results of the linear mixed-effects model analyses. In terms of form recall knowledge, depending on the scoring, the model explained from 69% to 70% of the variance in academic collocation knowledge due to both fixed and random effects (conditional  $R^2_{recall\ strict} = 0.70$ , conditional  $R^2_{recall\ lenient} = 0.69$ ). If only the fixed effects were considered, the model explained from 25% to 28% of the variance in the academic collocation knowledge (marginal  $R^2_{recall\ strict} = 0.28$ , marginal  $R^2_{recall\ lenient} = 0.25$ ). Among the fixed effects, regardless of the scoring schemes, similar results were found. Both knowledge of general high-frequency words (indicated by the CATSS3000 score) and academic words (indicated by the AVT score) were significantly related the knowledge of academic collocations (indicated by the ACT score) (all  $p$  values  $< 0.05$ ). However, knowledge of academic words seems to make a larger contribution than knowledge of general high-frequency words. In both scoring schemes, academic word knowledge had a medium positive effect on academic collocation knowledge ( $\beta_{recall\ strict} = 0.33$ ;  $\beta_{recall\ lenient} = 0.31$ ) while general word knowledge had a small positive effect ( $\beta_{recall\ strict} = 0.23$ ;  $\beta_{recall\ lenient} = 0.19$ ). In terms of form recognition knowledge, together the fixed and random effects explained 39% of the variance in academic collocation knowledge (conditional  $R^2_{recognition} = 0.39$ ) while the fixed effects explained 24% (marginal  $R^2_{recognition} = 0.24$ ). Among the fixed effects, knowledge of general high-frequency words did not significantly contribute to knowledge of academic collocation ( $p > 0.05$ ), while knowledge of academic words did ( $p < 0.05$ ). In particular, academic words had a medium positive effect on academic collocation knowledge ( $\beta_{recognition}$

= 0.30). In all cases, neither the length of study nor university entrance exam significantly affected academic collocation knowledge (all  $p$  values > 0.05).

[INSERT TABLE 6 HERE]

**TABLE 6.** Linear mixed-effects model with the ACT score as the dependent variable

Fixed effects	Recall (strict)			Recall (lenient)			Recognition		
	$\beta$	95% CI	<i>p</i>	$\beta$	95% CI	<i>p</i>	$\beta$	95% CI	<i>p</i>
Intercept	-0.01	[-0.56, 0.55]	0.08	-0.01	[-0.57, 0.55]	0.14	0.04	[-0.32, 0.40]	0.43
CATSS3000 score	0.23	[ 0.06, 0.40]	0.01	0.19	[ 0.03, 0.36]	0.02	0.14	[-0.02, 0.30]	0.08
AVT score	0.33	[ 0.17, 0.49]	<.001	0.31	[ 0.16, 0.47]	<.001	0.30	[ 0.10, 0.51]	0.004
Length of studying English	-0.03	[-0.15, 0.09]	0.59	-0.03	[-0.15, 0.09]	0.66	0.13	[-0.03, 0.30]	0.11
University Entrance Exam score	-0.09	[-0.20, 0.02]	0.09	-0.07	[-0.18, 0.03]	0.18	0.07	[-0.12, 0.26]	0.45

## DISCUSSION

The present study is among the few studies that have measured L2 learners' knowledge of academic collocations and is the first to examine the relationships between this knowledge and knowledge of academic words and general high-frequency words. By adapting parallel versions of recently developed tests to measure both form recall and form recognition knowledge and carefully tailoring them for the target participants, it provides further insights into L2 learners' knowledge of academic collocations and its relationships with knowledge of different kinds of single words.

The first research question is about L2 learners' knowledge of academic collocations. This study found that the participants scored 28.45% on the Recognition ACT and 47.12% on the Recall ACT. Regardless of whether form recall or form recognition was being measured, the participants in our study had lower knowledge of academic collocations than Nguyen's (2023) (31.58% and 64.7%). As the same collocations were measured in both studies, this finding could probably be because the present study measured knowledge of both English-major and non-English major students while Nguyen only focused on English majors.

Compared to non-English majors, English-major students may have more exposure to and practice with academic vocabulary. The academic collocational knowledge of the participants in the current study was also lower than that of Nguyen et al.'s (2024) and Lee's (2024) participants. Although the collocations tested in our study are different from those in Lee's study, they were the same as those in Nguyen et al.'s (2024). Therefore, one possible reason for the different findings could be the participants in our study were EFL learners who had much more limited exposure to English than all or a reasonable number of participants in the other studies, who were ESL learners. The differences in educational systems may also impact L2 learners' exposure to and practice with academic vocabulary.

The present study found that, although the participants in the recall and recognition groups were comparable in terms of the length of formal English instruction, university entrance exam score and program of study, the recall group always had lower scores in the academic collocation test than the recognition group. To some extent this finding is in line with those from previous studies (Lee, 2024; Nguyen et al., 2024) that L2 learners' form recall knowledge of academic collocations was smaller than their form recognition knowledge. Together the present study and earlier studies provide further evidence supporting González-Fernández and Schmitt's (2020) proposal that the key aspect in conceptualizing vocabulary knowledge is the recognition versus recall distinction.

The second research question concerned the relationships between L2 learners' knowledge of academic collocations and their knowledge of two kinds of single words. This study found that, regardless of the scoring scheme and level of vocabulary knowledge being measured, knowledge of academic words was always positively related to knowledge of academic collocations. In contrast, knowledge of general high-frequency words related to knowledge of academic collocations at the form recall level, but not at the form recognition level. To some extent this study supports previous studies (e.g., Nguyen, 2023) that knowledge of collocations is associated with knowledge of single words. However, the original contribution of this study lies in the respective relationships of knowledge of the two kinds of single words.

To begin with, it found that knowledge of academic words significantly contributed to knowledge of academic collocations at both the form recall and form recognition levels. Moreover, compared to general high-frequency words, knowledge of academic words always made a greater contribution. The greater contribution of academic words than general high-frequency words could be explained from both theoretical and pedagogical perspectives. According to connectionist models (e.g., Gasser, 1990), language is represented in the mind

as a network of multiple units. Some units are strongly associated with each other such that the presence of one may activate knowledge of the others. In contrast, other units are weakly associated with each other and the presence of one may hinder the activation of the others. On the other hand, usage-based approaches (e.g., Ellis, 2002; Tomasello, 2003) propose that learners encounter linguistic patterns in the input and the more frequently they encounter these patterns, the more likely it is that they connect them with each other and with the context in which they occur. In the light of these theories, it could be that academic collocations, academic words, and general high-frequency words relate to each other in learners' minds, but with different degrees of association. Similar to academic collocations, academic words occur frequently in academic contexts. Learners may see that academic words and academic collocations repeatedly appear together and develop a strong association between knowledge of these two kinds of vocabulary. In contrast, general high-frequency words occur frequently in a wide range of contexts, not just in academic text, which may make their connection with academic words weaker.

In addition to theory-related reasons, the stronger relationships of academic words than general high-frequency words could be explained from a pedagogical perspective. Different kinds of vocabulary may relate to different learning strategies (Gu, 2020; Nation, 2022).

When learning academic words, learners have developed specific strategies to learn academic vocabulary which may be transferred to their learning of academic collocations. In contrast, strategies developed for learning general high-frequency words may not be applicable to learning academic collocations.

In addition to finding the stronger relationship with academic words than general-high frequency words, this study also revealed that knowledge of the general words only significantly related to knowledge of academic collocations at the form recall level, but not at the form recognition level. The significant relationship at the form recall could be explained

as follows. Text comprehension tends to increase with the percentage of known words in the text (e.g., Schmitt et al., 2011). As general high-frequency words made up a reasonable proportion of a wide range of texts, including academic texts (Dang, 2020; Dang & Webb, 2020), having a larger knowledge of these words means that learners are more likely to comprehend the message in the input and have more cognitive resources to notice and learn novel vocabulary in the input, including academic collocations. This may explain the significant relationship between form recall knowledge of academic words and academic collocations found in the present study.

On the other hand, the lack of a significant relationship between knowledge of general high-frequency words and academic collocations at the form recognition level could be because, compared to form recall, form recognition knowledge is less demanding to acquire (Aviad-Levitzky et al., 2019; Laufer et al., 2004), and thus requires fewer cognitive resources to process. According to the depth of processing theory ( Craik & Lockhart, 1972), the deeper learners process the information, the better it will be retained in their minds. The fact that general high-frequency words also occur frequently in other contexts, not just academic contexts, and that form recognition knowledge requires a lighter level of processing than form recall means that learners will be less likely to remember the academic contexts in which they encounter general high-frequency words and academic collocations together. Consequently, at the form recognition level the link between knowledge of general high-frequency words and academic collocations may not be strong enough for a significant relationship to be evident.

Together the findings of the present study expand our understanding of the nature of vocabulary knowledge beyond the current conception of individual word knowledge as represented by Nation's framework. First, the inter-relation between academic collocations and single words found in this study indicates that collocation knowledge remains closely

related to knowledge of single words, which supports Nation's framework. However, the fact that only weak to moderate effect sizes were found suggests that collocation knowledge may include unique elements that are not captured by knowledge of single words alone. As Nation's framework currently treats collocational knowledge as part of single word knowledge without distinguishing how different kinds of vocabulary influence collocation knowledge, our finding suggests the need for a broader theoretical approach to capture knowledge of multiword sequences, including collocations. This study then provides empirical evidence supporting Nation and Webb's (2011) framework of multiword sequences. It also expands on Nation and Webb's framework, especially the association aspect. The stronger relationship of academic words with academic collocations compared to general high-frequency words indicates that collocational knowledge is not always uniformly tied to general vocabulary knowledge. Instead, it is sensitive to domain-specific vocabulary knowledge.

In addition to expanding current literature on L2 learners' vocabulary knowledge, the present study also contributes to vocabulary research methodologically. In particular, the preliminary study involved a more thorough evaluation of the proposed tests than is typically undertaken in vocabulary research studies, with particular consideration given to the suitability of the instruments for the target population of university students in Vietnam. This included statistical analyses, expert judgements, and feedback from a sample of test-takers. One result was that ACT Recognition, a previously published test, was completely rewritten to make it a more valid measure of the ability to recognise academic collocations. Significant revisions were made to the other tests as well to improve their quality as instruments for use in the main study.

Though not the main focus of the present study, several findings related to the effects of the length of formal English instruction and entry language proficiency levels on academic

vocabulary knowledge have emerged. The length of studying English did not significantly contribute to the participants' knowledge of academic collocations. This non-significant effect could be because the students only start learning academic vocabulary when entering their university. Therefore, their pre-university studies may not have a clear impact on their academic collocation knowledge. In addition, the participants' scores on the university entrance exam did not significantly contribute to their knowledge of academic collocations. This finding could be because pre-university education in Vietnam focused more on developing knowledge of single words rather than collocations (Vu & Peters, 2021).

The present study has several limitations. It only focused on Vietnamese EFL learners. Replication studies with learners in other contexts would be useful. The second limitation is that this study employed off-line measures to assess learners' vocabulary knowledge. Future research combining these measures with on-line measures such as event-related brain potential (ERP) and follow-up interviews with participants would provide further insights into the participants' knowledge of academic collocations, academic words, and general words and their relationships. The third limitation is that this study did not measure knowledge of general collocations. There are several reasons for this decision. The primary focus of our project was knowledge of academic collocations and its relationships with different kinds of single words (academic words and general high-frequency words). Such findings would help to expand our understanding of the nature of vocabulary knowledge beyond the current conception of individual word knowledge, as represented by Nation's (2022) framework, by connecting knowledge of multiword units with that of single words in academic contexts. Moreover, our study aimed to measure vocabulary knowledge at both recall and recognition levels. To achieve this aim, we needed to have pairs of parallel recall and recognitions versions. However, no suitable tests of general collocations were available. Additionally, there is the constraint of avoiding too much testing of individual students. In

our study, each learner had to complete an academic collocation test, an academic word test, a general word test, and a language learning experience questionnaire. Asking them to do another test would have caused fatigue effects and negatively impacted on the findings of our study. However, future research could consider measuring knowledge of general collocations to provide better insights into the relationships of different kinds of vocabulary knowledge.

The fourth limitation of our study is that the recognition version of the Academic Vocabulary Test is in the matching format, which is slightly different from the formats of the ACT and CATSS. However, we chose this test because it is the only test of academic words with parallel versions that have gone through a rigorous validation process. Moreover, to a certain extent, the Academic Vocabulary Test also measures knowledge of form recognition because test takers have to select among the given options the word form that best matches the provided meaning. Also, even if the matching format measures knowledge of meaning recognition, previous studies (Aviad-Levitzky et al., 2019; Laufer et al. 2004) have found that there was no significant difference in the degree of difficulty between form recognition and meaning recognition. Therefore, it is unlikely that that knowledge will seriously impact on the results of our study.

## **PEDAGOGICAL IMPLICATIONS**

This study found that knowledge of academic collocations was positively related to knowledge of academic words at both the form recall and form recognition levels and to knowledge of general high-frequency words at the form recall level. This finding indicates that knowledge of both kinds of single words is important for the development of knowledge of academic collocations. As knowledge of academic words had a stronger relationship with knowledge of academic collocations than knowledge of general high-frequency words, EAP programs should prioritize teaching academic words over general high-frequency words.

They should offer opportunities for learners to encounter academic words and academic collocations together in academic discourse to strengthen the connections between these two groups of vocabulary items, which may then facilitate the learning of academic collocations. However, it does not mean that we should dismiss the value of general high-frequency words. The positive relations between knowledge of general high-frequency words and academic collocations at the form recall level indicates that learners with a large knowledge of general high-frequency words are likely to have a larger knowledge of academic collocations. Given that a reasonable number of learners have insufficient knowledge of general high-frequency words when starting EAP studies (e.g., Dang, 2020; Lu & Dang, 2023), pre-university education should pay more attention to general high-frequency words, so that students can have a solid knowledge of these words prior to their university study. Knowledge of these words not only enables learners to communicate in a wide range of contexts but also creates a solid foundation for their learning of academic collocations. Moreover, this study found that the positive relationship between knowledge of general high-frequency words and academic collocation existed in the case of form recall but not form recognition. This indicates that when teaching general high-frequency words, teachers should help students to reach the higher level of knowledge represented by form recall. This may then reinforce the association between general high-frequency words and academic collocations when they are encountered together in academic contexts, which may facilitate the learning of academic collocations. Finally, the participants in the present study had a modest knowledge of academic collocations compared to EFL learners in other contexts. This finding indicates that more guidance should be provided to support these students in broadening their knowledge of academic collocations. The support should help to develop both recognition and recall knowledge given the participants' smaller recall knowledge in both kinds of vocabulary.

## CONCLUSION

This study is the first to systematically examine the relationships between EFL learners' knowledge of academic collocations and their knowledge of academic words and general high-frequency words. It found that in the case of form recall, both knowledge of academic words and general high-frequency words positively associated with knowledge of academic collocations, with academic words having a stronger relationship. However, in the case of form recognition, only knowledge of academic words had a positive relationship with knowledge of academic collocations. These findings indicate the relative value of knowledge of academic words and general high-frequency words for the learning of academic collocations. They also indicate the need for a broader theoretical approach to capture knowledge of multiword sequences, especially in specialised contexts, and its relationships with knowledge of single words.

## NOTES

<sup>1</sup> These guidelines took account of test-taking strategies that Vietnamese students are commonly trained to apply in preparing to answer multiple-choice items in university entrance exams.

<sup>2</sup> Following Nguyen et al. (2024), we examined four properties of the tests through Rasch analysis: item measure, item fit, unidimensionality, and local independence, and these assumptions were met.

<sup>3</sup> We did not use L1 translation for the stem sentences, multiple-choice options, and definition clues of the tests for several reasons. First, although these sentences, options, and clues are in L2, they were extensively piloted in our preliminary study to ensure they were relevant to our target participants. Moreover, the recognition tests focused on form recognition, making L1 translation of the multiple-choice options inappropriate. Also, the original test pairs, which

have undergone rigorous validation by their developers, are monolingual. Translating the stem sentences, definition clues, and options into Vietnamese would have required additional validation to ensure the accuracy of the L1 versions. From a practical standpoint, this would be both time- and resource-intensive, and our pilot suggests it may not be necessary given that the monolingual tests worked well with our participants.

<sup>4</sup> Analysis using the Kolmogorov-Smirnov and the Shapiro-Wilk test both indicated that the data were not normally distributed for length of study ( $p < 0.001$ ) and university entrance exam score ( $p < 0.001$ ).

<sup>5</sup> This abbreviation refers to CATSS scores on the most frequent 3,000 words of general English (general high-frequency words)

## REFERENCES

Ackermann, K., & Chen, Y. H. (2013). Developing the Academic Collocation List (ACL) – A corpus driven and expert-judged approach. *Journal of English for Academic Purposes*, 12(4), 235–247. <https://doi.org/10.1016/j.jeap.2013.08.002>

Aviad-Levitzky, T., Laufer, B., & Goldstein, Z. (2019). The New Computer Adaptive Test of Size and Strength (CATSS): Development and validation. *Language Assessment Quarterly*, 16(3), 345–368. <https://doi.org/10.1080/15434303.2019.1649409>

Carr, N. T. (2011). *Designing and analyzing language tests*. Oxford University Press

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Taylor & Francis.

Craik, F., & Lockhart, R. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11(6), 671–684.

Crossley, A. S., Salsbury, T., & McNamara, D. S. (2015). Assessing lexical proficiency using analytic ratings: A case for collocation accuracy. *Applied Linguistics*, 36(5), 570–590.

<https://doi.org/10.1093/applin/amt056>

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. (2020). Vocabulary and good language teachers. In C. Griffiths & Z. Tajeddin (Eds.), *Lessons from good language teachers* (pp. 203–218). Cambridge University Press.

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

Ellis, N. C., Simpson-Vlach, R., & Maynard, C. (2008). *TESOL Quarterly*, 42(3), 375–396. <https://doi.org/10.1002/j.1545-7249.2008.tb00137.x>

Gardner, D., & Davies, M. (2014). A new academic vocabulary list. *Applied Linguistics*, 35(3), 305–327. <https://doi.org/10.1093/applin/amt015>

Gasser, M. (1990). Connectionism and universals of second language acquisition. *Studies in Second Language Acquisition*, 12(2), 179–199. <https://doi.org/10.1017/S0272263100009074>

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, P. (2020). Strategies for learning vocabulary. In *The Routledge handbook of vocabulary studies* (pp. 271–287). Routledge.

Kane, M. (2013). The argument-based approach to validation. *School Psychology Review*, 42(4), 448–457. <https://doi.org/10.1080/02796015.2013.12087465>

Laufer, B., Elder, C., Hill, K., & Congdon, P. (2004). Size and strength: Do we need both to measure vocabulary knowledge? *Language Testing*, 21(2), 202–226.

<https://doi.org/10.1191/0265532204lt277oa>

Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning*, 54(3), 399–436. <https://doi.org/10.1111/j.0023-8333.2004.00260.x>

Lee, S. (2024). The relationship between receptive and productive knowledge of L2 English collocations. *International Journal of Applied Linguistics*, 1–25.

<https://doi.org/10.1111/ijal.12605>

Lu, C., & 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*. <https://doi.org/10.1177/13621688231155820>

Martinez, R., & Murphy, V. A. (2011). Effect of frequency and idiomaticity on second language reading comprehension. *TESOL Quarterly*, 45(2), 267–290.

<https://doi.org/10.5054/tq.2011.247708>

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

Nation, I. S. P. (2001). *Learning vocabulary in another language* (1<sup>st</sup> ed.). Cambridge University Press. <https://doi.org/10.1017/9781009093873>

Nation, I. S. P. (2022). *Learning vocabulary in another language* (3rd ed.). Cambridge University Press. <https://doi.org/10.1017/9781009093873>

Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, 31(7), 9–13. <https://doi.org/10.26686/wgtn.12552197.v1>

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

Nguyen, T. M. H. (2023). Vietnamese English majors' knowledge of academic collocations and its relationship with vocabulary size. *3rd VNZ TESOL International Conference 2023 Proceeding*, 193–203.

Nguyen, T. M. H., Coxhead, A., & Gu, P. (2024). Argument-based validation of Academic Collocation Tests. *Language Testing*, 41(3), 459–505.

<https://doi.org/10.1177/026553222311984>

Nguyen, T. M. H., & Webb, S. (2017). Examining second language receptive knowledge of collocation and factors that affect learning. *Language Teaching Research*, 31(3), 98–320.

Öksüz, D., Brezina, V., & Rebuschat, P. (2021). Collocational processing in L1 and L2: The effects of word frequency, collocational frequency, and association. *Language Learning*, 71(1), 55–98. <https://doi.org/10.1111/lang.12427>

Pecorari, D., Malmström, H., & Warnby, M. (2025). A test of controlled productive knowledge of English academic vocabulary. *Acta Didactica Norden*, 19(1), 1–32.

<https://doi.org/10.5617/adno.11584>

Pecorari, D., Shaw, P., & Malmstrom, H. (2019). Developing a new academic vocabulary test. *Journal of English for Academic Purposes*, 39, 59–71.

<https://doi.org/10.1016/j.jeap.2019.02.004>

Plonsky, L., & Oswald, F. L. (2014). How big is “big”? Interpreting effect sizes in L2 research. *Language Learning*, 64(4), 878–912. <https://doi.org/10.1111/lang.12079>

- Puimège, E., & Peters, E. (2020). Learning formulaic sequences through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition*, 42(3), 525–549.
- Read, J. (2000). *Assessing vocabulary*. Cambridge University Press.  
<https://doi.org/10.1017/cbo9780511732942>
- Read, J. (2023). Towards a new sophistication in vocabulary assessment. *Language Testing*, 40(1), 40-46. <https://doi.org/10.1177/02655322221125698>
- Read, J., & Dang, T. N. Y. (2022). Measuring depth of academic vocabulary knowledge. *Language Teaching Research*. <https://doi.org/10.1177/13621688221105913>
- Schmitt, N. (2023). Norbert Schmitt's essential bookshelf: Formulaic language. *Language Teaching*, 56(3), 420–431. <https://doi.org/10.1017/s0261444822000039>
- Schmitt, N., Nation, P., & Kremmel, B. (2020). Moving the field of vocabulary assessment forward: The need for more rigorous test development and validation. *Language Teaching*, 53(1), 109-120. <https://doi.org/10.1017/S0261444819000326>
- 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.  
<https://doi.org/10.1111/j.1540-4781.2011.01146.x>
- Sonbul, S., El-Dakhs, D. A. S., & Masrai, A. (2023). Second language productive knowledge of collocations: Does knowledge of individual words matter? *Studies in Second Language Acquisition*, 45(2), 480–502. <https://doi.org/10.1017/s0272263122000341>
- Stoeckel, T., McLean, S., & Nation, P. (2021). Limitations of size and levels tests of written receptive vocabulary knowledge. *Studies in Second Language Acquisition*, 43(1), 181-203.  
<https://doi.org/10.1017/S027226312000025X>

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Harvard University Press.

Vilkaitė, L. (2017). Incidental acquisition of collocations in L2: Effects of adjacency and prior vocabulary knowledge. *ITL - International Journal of Applied Linguistics*, 168(2), 248–277. <https://doi.org/10.1075/itl.17005.vil>

Vu, D. V., & Peters, E. (2021a). Vocabulary in English language learning, teaching, and testing in Vietnam: A review. *Education Sciences*, 11(9), 563. <https://doi.org/10.3390/educsci11090563>

Vu, D. V., & Peters, E. (2021b). Incidental learning of collocations from meaningful input: A longitudinal study on three reading modes and factors that affect learning. *Studies in Second Language Acquisition*, 44(3), 1–23.

Webb, S. (2021). A different perspective on the limitations of size and level tests of written receptive vocabulary knowledge. *Studies in Second Language Acquisition*, 43(2), 454–461. <https://doi.org/10.1017/s0272263121000449>

Webb, S., Sasao, Y., & Balance, O. (2017). The Updated Vocabulary Levels Test: Developing and validating two new forms of the VLT. *ITL-International Journal of Applied Linguistics*, 168(1), 34–70. <https://doi.org/10.1075/itl.168.1.02web>