

# Vocabulary Size and Depth of Knowledge: A Study of Bahraini EFL Learners

Adel W. Alqallaf<sup>1</sup> & Madani O. Ahmed<sup>1</sup>

<sup>1</sup> Department of English Language, University of Bahrain, Sakhir, Bahrain

Correspondence: Adel Alqallaf, Department of English Language, University of Bahrain, Sakhir, Bahrain.  
E-mail: a.waelq@live.com

Received: October 5, 2021      Accepted: November 8, 2021      Online Published: November 19, 2021

doi:10.5539/ijel.v12n1p58      URL: <https://doi.org/10.5539/ijel.v12n1p58>

## Abstract

This study investigates the size and depth of vocabulary knowledge and its relationship to the general language proficiency of EFL learners. The study sample included 120 students from the University of Bahrain. The sample was randomly selected from the student population and split into two groups in terms of their level of English: intermediate and advanced. The study aims to answer four questions: (1) What is the effect of general language proficiency on the sizes of the receptive and productive vocabularies of learners of English at the University of Bahrain? (2) How does general language proficiency affect the depth of vocabulary knowledge of learners of English at the University of Bahrain? (3) What is the relationship between receptive and productive vocabularies and the depth of vocabulary knowledge? and (4) What is the relationship between vocabulary size and the nature of lexical networking? All the students in the sample completed three vocabulary tasks. The first two tasks were Meara and Jones's Eurocentres Vocabulary Size Test (1990) and Meara and Fitzpatrick's Lex30 word association task (2000), which were used to measure the sizes of receptive and productive vocabularies. The third task was Gyllstad's COLLEX test (2007), which was used to investigate the depth of vocabulary knowledge. A quasi-experimental approach was adopted using a quantitative approach to analyze the data. The data of the study were analyzed by comparing the results of the two groups in relation to the three tasks using SPSS 16.0. The findings of the study have revealed that general language proficiency has a positive effect on learners' receptive vocabulary size, a moderate effect on learners' productive vocabulary size, and a very low effect on the depth of vocabulary knowledge. In addition, no relationship was shown between the size of vocabulary and the nature of lexical networking. With reference to these results, pedagogical and future research recommendations are made.

**Keywords:** vocabulary knowledge, vocabulary size, vocabulary depth, productive vocabulary, receptive vocabulary, vocabulary measurement

## 1 Introduction

It is widely accepted that vocabulary plays a significant role in the acquisition of English as a second language (Knight, 1994; Laufer, 1997; Yoshii & Flaitz, 2013). Vocabulary knowledge is known to be one of the fundamental components in language acquisition (Schmitt, 2008), and it has been considered a major factor in language proficiency (Alavi & Akbarian, 2012; Read, 2004). As Richards and Rodgers (2001, p. 132) strongly assert, "the building blocks of language learning and communication are not grammar, function, notions, or some other unit of planning and teaching but lexis, that is, word and word combinations." To reinforce this, Namaziandost, Kumar and Çakmak (2021, p. 2) mention that vocabulary is "regarded as mandatory to acquiring knowledge through receptive skills, reading and listening, and essential to the development of the productive skills of communicating and writing."

Wilkins (1972, p. 111) argues that "without grammar very little can be conveyed, without vocabulary nothing can be conveyed." It is impossible to reach a mutual understanding of the context with inaccurate vocabulary, such as when describing a car using the word *sky* instead, as in *he has a red sky*. Thus, in order to master an L2 (second language), learners and language instructors must pay more attention to vocabulary knowledge.

Recent literature has recognized two substantial aspects related to vocabulary knowledge (Schmitt, 2014). The first aspect is known as size, or breadth, of vocabulary knowledge, which can be defined as how many words learners know. The second is known as depth of vocabulary knowledge; that is, to what extent learners know

those words. On the one hand, vocabulary size refers to the linear and one-dimensional domains of words. On the other, depth of vocabulary knowledge is related to linguistic characteristics. These characteristics include the meaning of words, semantic relationships, collocations, and syntactic patterns (Bardakçi, 2016).

It is evident that the size and depth of vocabulary knowledge are not difficult to measure (Meara, 2009; Milton, 2009). First, the size of vocabulary knowledge can be evaluated by measuring two types of vocabulary: receptive and productive. Second, the depth of vocabulary knowledge can be investigated through assessing the semantic relationships of words. This goes hand in hand with Schmitt's (2014) argument that vocabulary depth can be assessed by focusing on word associations and collocations.

Researchers nowadays have a prominent interest in vocabulary knowledge. Therefore, literature shows that some of them link this topic to the four skills of the language learning, listening, reading, writing, and speaking, and some of them link the topic to other linguistic aspects (Luo, Wan, & Zhang, 2021; Namaziandost, Kumar, & Çakmak, 2021; Wero, Machmud, & Husain, 2021; Kalay, 2021; Ha, 2021; Enayat & Derakhshan, 2021).

This study investigates the role of the size and depth of vocabulary knowledge in the language acquisition journey made by learners of EFL (English as a foreign language) at the University of Bahrain. These two interrelated topics play a significant role in the development of vocabulary knowledge, which is one of the essential elements in L2 acquisition.

This study draws on two complementary and overlapping theoretical frameworks of vocabulary knowledge. While the first is known as Richards's (1976) vocabulary knowledge framework, the second is known as Nation's (2001) vocabulary knowledge framework. The combination of these two frameworks embraces a sense of fully-fledged structure for comprehending the concept of vocabulary knowledge.

### 1.1 Richards's (1976) Vocabulary Knowledge Framework

This framework describes the nature of knowing a word as a multi-dimensional domain. Vocabulary knowledge is not only about knowing the meaning of words; it is also related to several linguistic aspects, such as semantic, syntactic, and paradigmatic features. Eight assumptions are provided in this framework in an attempt to describe this multi-dimensional domain, in which assumptions two to eight define the meaning of knowing a word, as shown below (Richards, 1976, pp. 79–83):

“Knowing a word means:

- 1) the native speaker of a language continues to expand their vocabulary in adulthood, whereas there is comparatively little development of syntax in adult life.
- 2) knowing the degree of probability of encountering that word in speech or print. For many words, we also know the sort of words most likely to be found associated with that word.
- 3) knowing the limitations imposed on the use of the word according to variations of function and situation.
- 4) knowing the syntactic behavior associated with that word.
- 5) knowing the underlying form of the word and the derivatives that can be made from it.
- 6) knowing the network of associations between that word and other words in language.
- 7) knowing the semantic value of the word.
- 8) knowing many of the different meanings associated with that word.”

Richards defines the meaning of vocabulary knowledge through seven assumptions (numbered two to eight). Assumption two suggests that the frequency of words has a significant role in communication. Some words are used more frequently than others. For example, a word like *day* can be used in more contexts in comparison to words like *nook* or *swath*. In addition, assumption two asserts that knowing a word requires the learners to know the associations of that word. For instance, when mentioning the word *blood*, there is a probability that words like *bank* and *donation* will be brought to mind. The third assumption argues that some words are specified to be used in specific fields. For example, words like *agent*, *ally*, *embassy*, and *protocol* are used in diplomatic fields, whereas *diagnosis*, *patient*, *injury*, and *procedure* are used in medical fields. The fourth assumption proposes that knowing a word means knowing its collocations. The fifth assumption suggests that knowing a word means to also know the derivative forms of the base word. For example, knowing the word *connect* means to also know the words *connected*, *disconnected*, and *reconnect*. The sixth assumption argues that knowing a word means to be aware of the word's associations, including synonyms, antonyms, and subordinate classification. For example, to know the word *star* is to know associated words such as *universe*, *moon*, *shine*, and *bright*. Richards's seventh assumption suggests that knowing a word is to know that there are different values of the word semantically:

concrete words like *cup*, *student*, *table*, and *chair* have more semantic value in comparison to words like *peace*, *pain*, *love*, and *feel*, which are known as abstract words. Finally, the last assumption of Richards's framework proposes that knowing a word is to know its different and multiple meanings. For example, the word *bow* can mean a knot tied with two loops and two loose ends, or bending as a sign of respect.

There has been criticism in the literature that Richards's framework is not an integrated framework of vocabulary knowledge. Besides the lack of convention in the productive and receptive vocabularies, Meara (1996) comments that Richards's assumptions cannot be held as a systematic account of word knowledge for several reasons. One of these is that Richards's endeavor to identify word knowledge is related to another genre, in which it is "an honest attempt to give an account of contemporary linguistic research with inferences and applications to teaching where appropriate" (Meara, 1996, p. 2). Added to that, for example, assumption four is mainly involved in a "short-lived development in syntactic theory, case grammar, which faded shortly" after the appearance of these assumptions (Meara, 1996, p. 2). Assumption eight "does not appear to be based on any specific research" but, instead, is based on "language teachers' points of view", which show a clear gap in the research at that time (Meara, 1996, p. 2). By contrast, Nation's (2001) recent contribution to the literature provides an improved framework of vocabulary knowledge, which is discussed in the next section.

### 1.2 Nation's (2001) Vocabulary Knowledge Framework

The second theoretical framework of this research is known as Nation's vocabulary knowledge framework. According to Schmitt (2010), this framework has been regarded as one of the most widely used by researchers. Nation states that vocabulary knowledge is composed of three categories. As Table 1 demonstrates, these categories are *form*, *meaning*, and *use*, which are essentially related to both receptive (*R*) and productive (*P*) aspects of word knowledge.

Table 1. What is involved in knowing a word (Nation, 2001, p. 27)

<b>F</b>	<i>spoken</i>	<i>R</i>	<i>What does the word sound like?</i>
<b>o</b>		<i>P</i>	<i>How is the word pronounced?</i>
<b>r</b>	<i>written</i>	<i>R</i>	<i>What does the word look like?</i>
<b>m</b>		<i>P</i>	<i>How is the word written and spelled?</i>
	<i>word parts</i>	<i>R</i>	<i>What parts are recognizable in this word?</i>
		<i>P</i>	<i>What word parts are needed to express this meaning?</i>
<b>M</b>	<i>form and meaning</i>	<i>R</i>	<i>What meaning does this word form signal?</i>
<b>e</b>		<i>P</i>	<i>What word form can be used to express this meaning?</i>
<b>a</b>	<i>concept and referents</i>	<i>R</i>	<i>What is included in the concept?</i>
<b>n</b>		<i>P</i>	<i>What items can the concept refer to?</i>
<b>i</b>	<i>associations</i>	<i>R</i>	<i>What other words does this make us think of?</i>
<b>n</b>		<i>P</i>	<i>What other words could we use instead of this one?</i>
<b>g</b>			
<b>U</b>	<i>grammatical functions</i>	<i>R</i>	<i>In what patterns does the word occur?</i>
<b>s</b>		<i>P</i>	<i>In what patterns must we use this word?</i>
<b>e</b>	<i>collocations</i>	<i>R</i>	<i>What words or types of words occur with this one?</i>
		<i>P</i>	<i>What words or types of words must we use with this one?</i>
	<i>constraints on use (register, frequency...)</i>	<i>R</i>	<i>Where, when, and how often would we expect to meet this word?</i>
		<i>P</i>	<i>Where, when, and how often can we use this word?</i>

As indicated above, Nation's framework shows three categories of vocabulary knowledge. These categories encompass nine degrees of knowledge. First, there is the form category, which includes knowledge of the spoken form, written form, and parts of words. Second, the meaning category shows knowledge of the form and meaning, concept and referents, and associations. The third is the use category, which includes knowledge of the grammatical functions, collocations, and constraints on use.

Thus, knowing a word form embraces the spoken form, written form, and word parts. Both spoken and written forms assist learners of the language to enhance their lexical knowledge by improving their literacy. Added to this, knowledge of phonics, word recognition, and spelling provides L2 learners with an advantage in recognizing and decoding word meaning. Knowing the spoken form refers to the receptive ability to understand and comprehend words when they are heard. In addition, knowing the spoken form refers to the productive ability of pronouncing words correctly. Furthermore, as a receptive domain, knowing the written form refers to the ability of recognizing words when read, and, as a productive domain, knowing the written form refers to the

knowledge of spelling words correctly (Schmitt, 2000).

Prefixes, suffixes, and stems are involved when discussing the knowledge of word parts. It is possible to know the meaning of any unknown word when the prefix, suffix, or stem of the word is known. For example, when learners know that the word *unbelievable* contains the prefix *un-*, which means *not*, and the suffix *-able*, which means *can be*, then it is easy to know that the meaning of the word is *not to be believed*. Also, Nation believes that the connection between form and meaning, concept and referents, and word associations is meant to be the knowledge of word meaning. For example, when a word is seen or heard, the meaning of this word is retrieved simultaneously; similarly, when a meaning is needed to be expressed, the form of the word also comes to mind. As Richards (1976, p. 81) claims, words are not stored separately in blocks in a human being's mind: "words do not exist in isolation." Words are linked or joined together with their associations in the minds of human beings. Each single word is stored in the mental lexicon of language learners with its associations—words that are related to each other and share common things. For example, words like *school*, *chair*, *table*, *classroom*, *students*, and *teachers* can be stored together in the mental lexicon.

In addition, Nation argues that the knowledge of using a word correctly refers to the grammatical function of that word, collocations, and an awareness of the constraints of that word. To make it clear, the grammatical function factor plays a vital role when using a word. For example, learners do not utter sentences like *I reading a lot* or *I eaten a lot*; instead, they say *I read a lot* and *I eat a lot*. Choosing correct word classes is needed to form grammatical sentences. Added to that, register and frequency are two other factors that are essential in using words. Schmitt (2000, p. 31) defines register as the stylistic constraints that "make each word more or less appropriate for certain language situations or language purposes." Frequency, however, is a different word capacity level; e.g., a word like *laugh* is used more than *guffaw*, *giggle*, and *chuckle*, and therefore it is a high frequency word. That is, the process of recalling and recognizing high frequency words is easy to do in the mental lexicon. Hence, learners have to be aware of how to use the word with its constraints.

### 1.3 Research Problem

Literature shows that there is a strong emphasis on assessing vocabulary knowledge and its importance in achieving excellence in learning EFL (Meara, 2009; Milton, 2009; Schmitt, 2010; Nation & Webb, 2011). A systematic study examining the extent of the general language proficiency of EFL learners in terms of the size and depth of vocabulary knowledge in the context of Bahrain is lacking. Therefore, this study sheds light on the importance of measuring the size and depth of the vocabulary knowledge of EFL learners at the University of Bahrain (ranked number one in the Kingdom of Bahrain). This study investigates the extent to which the size and depth of vocabulary knowledge affect the general language proficiency of EFL learners who are taking English language courses at the University of Bahrain.

### 1.4 Research Objectives

To answer the research questions stated above, this study aims to do the following:

- 1) Measure the development of the receptive and productive vocabulary sizes of EFL learners at the University of Bahrain. In particular, the study focuses on the effect of general language proficiency on the development of both the size and depth of vocabulary knowledge, and the nature of mental lexical networking.
- 2) Evaluate the organization of the mental lexical networking of EFL learners at the University of Bahrain to examine the role of general language proficiency.
- 3) Investigate the relationship between the size of vocabulary and the organization of lexical networking in relation to general language proficiency.

### 1.5 Research Questions

The research questions for this study stem from the call in literature for more emphasis on measuring the size and depth of vocabulary knowledge. Several lines of evidence report on the importance of vocabulary for EFL learners. Therefore, this study aims to answer the following questions:

- 1) What is the effect of general language proficiency on the sizes of the receptive and productive vocabularies of learners of English at the University of Bahrain?
- 2) How does general language proficiency affect the depth of vocabulary knowledge of learners of English at the University of Bahrain?
- 3) What is the relationship between receptive and productive vocabularies and the depth of vocabulary knowledge?

4) What is the relationship between vocabulary size and the nature of lexical networking?

## 2. Method

Measuring the size and depth of vocabulary knowledge requires dealing with scores. Scores enable the researcher to come up with valuable recommendations by analyzing statistical findings. Therefore, this research was conducted using a quantitative methodology. As mentioned before, this study investigates the size and depth of vocabulary knowledge and its relationship to the general language proficiency of EFL learners.

### 2.1 Population and Sample

The population of this study was University of Bahrain English major students at various proficiency levels. 120 L2 learners of English were chosen from the Department of English Language and Literature in the College of Arts. These participants, both male and female, were selected randomly based on their language proficiency level. They were then divided into two groups in terms of their proficiency level. As shown in Table 2, group one consists of 60 first-year intermediate students, aged between 18 and 19: 42 females and 18 males. Group two consists of 60 fourth-year advanced students, aged between 20 and 21: 45 females and 15 males. All participants in the study had started to learn English by the age of six or seven (Rixon, 2013), which generally refers to Grade 1. The proficiency levels of the two groups were represented by two English courses provided by the university: participants at the intermediate level were enrolled on the ENGL130 course (3 sections were involved), whereas participants at the advanced level were enrolled on the ENGL450 course (3 sections were involved).

Table 2. Participants in the study

Group	Language level	Number	Frequency	Valid percent
Group 1	Intermediate	60	Female: 42	35%
			Male: 18	15%
Group 2	Advanced	60	Female: 45	37.5%
			Male: 15	12.5%
<b>Total</b>		120	120	100%

### 2.2 Research Tools

This study uses three different tools to assess three aspects of vocabulary knowledge: receptive vocabulary, productive vocabulary, and vocabulary depth. Each aspect has its own measurement tool. The first tool is the EVST, used to measure receptive vocabulary knowledge (Meara & Jones, 1990). The second tool is the Lex30 word association task, which is used to measure the productive vocabulary knowledge (Fitzpatrick & Meara, 2004). The third tool of this research is the COLLEX test, used to investigate the depth of vocabulary knowledge (Gyllstad, 2007).

### 2.3 Eurocentres Vocabulary Size Test (EVST)

The EVST was developed by Meara and Jones (1990) to measure the size of learners' receptive vocabulary knowledge based on Thorndike and Lorge's (1944) most frequent 10,000 words of English. This test includes 300 words divided into five blocks. Each block represents 1,000 frequent words: "the first block contains a sample of items from the first 1,000 most frequent words in English; the second block is a representative sample from the second 1,000 most frequent words in English; and so on" (Meara, 1990, p. 3). Each block contains 60 words, although not all of them are real English words. These 60 words are divided into 40 real words and 20 non-words. In order to complete this test, the test takers need to mark the words that are familiar to them and leave the words that they do not know.

Historically, the format of the EVST was derived from one of the simple formats used to measure the size of receptive vocabulary knowledge, and it first appeared in the field of L1 studies (Sims, 1929; Tilley, 1936; Zimmerman, Broder, Shaughnessy, & Underwood, 1977). The format of this test is known as the 'checklist', which includes a list of words with the requirement for test takers to mark those familiar to them. However, later on, researchers realized that the result of this test could be affected by the overestimation of knowledge factor. Therefore, Anderson and Freebody (1983) found that the list of words in the EVST lacked pseudo-words (non-words) to maintain accurate results. Thus, they developed the test by adding 20 pseudo-words to the real words. These pseudo-words help to provide an accurate knowledge estimation of real words. Later on, Meara and Buxton (1987) used this developed EVST in the L2 field as a first attempt to check its workability. The test was then developed further by separating the words into 40 real words and 20 pseudo-words in each of the five levels, and a computer-based format was also provided (Meara & Jones, 1988, 1990).

A reliability test was run using the Statistical Package for the Social Sciences (SPSS), as can be seen in Table 3: the Cronbach's alpha is .887, which indicates a highly reliable test.

Table 3. EVST reliability

Cronbach's alpha	Cronbach's alpha based on standardized items	N. of items
.887	.763	6

### 2.3.1 Lex30 Word Association Task

In addition to the EVST, which is used to measure the size of receptive vocabulary knowledge, the study also measures the size of productive vocabulary knowledge using a method known as Lex30, a word association task designed by Meara and Fitzpatrick (2000). Lex30 contains a list of 30 stimulus words. To fulfil the task, test takers need to produce a range of one to four responses, or associations, to these stimuli (Appendix 2 contains a full version of Lex30). The following example shows one of the stimulus words, *animal*, with its possible associations of *elephant*, *tiger*, *farm*, and *wild*.

*animal: elephant – tiger – farm – wild*

As Meara (2009) argues, all the stimulus words provided in Lex30 were chosen in accordance with a number of criteria. The first criterion is that all of the 30 stimulus words are highly frequent; they were chosen from the first 1,000-word list by Nation (1984). Nation's 1,000-word list comprises low-level words, and learners are expected to identify and recognize them. The reason behind choosing these words is to make the task flexible so that it can be used with varying ranges of proficiency levels. The second criterion for choosing these words is that they do not elicit a single response, or association, like the words *black* and *dog*. To put it simply, words that may lead to a very narrow range of associations are not used in Lex30, whereas Meara and Fitzpatrick (2000) selected words that generate a wide range of associations. The third criterion is that each item on the stimulus words list naturally leads to associations that are not common words, and, by avoiding the usage of common words that are given by native speakers of English, learners who would like to measure their productive vocabulary size have the opportunity to generate a wide range of associations.

When Lex30 was developed, it was criticized for not having a clear vision of validity and reliability (Baba, 2002). However, in 2004, Fitzpatrick and Meara proved that Lex30 is valid and reliable by conducting a test-retest study and two concurrent measures of validity. One of these two measures used native speakers' data, and the other used two parallel tests: the productive version of the VLT (Laufer & Nation, 1995), and a translation task that required learners to translate contexts from Chinese to English, as Table 4 illustrates.

Table 4. Reliability and validity of Lex30 (Fitzpatrick & Meara, 2004, pp. 46–55)

#	Study	N	Main details
1	Reliability S.	16 L2 users of English	Lex30 has a high degree of reliability: - $0.866 (p < .01)$
2	Validity S. (N.1): (Native speakers' norms)	46 NS & 46 NNS	Native speakers produced a high rate of low-frequency words.
3	Validity S. (N.2): (Parallel tests)	55 EFL learners	A high correlation between the translation test, the VLT, and Lex30: - Lex30 & Productive Levels Test: $0.504 (p < .01)$ - Lex30 & Translation Test: $0.651 (p < .01)$ - Productive Levels Test & Translation Test: $0.843 (p < .01)$

Note. NS: native speakers; NNS: non-native speakers.

The reliability test-retest study consisted of 16 L2 users of English who were from different L1 backgrounds with different levels of language proficiency: lower intermediate to advanced. Participants completed the task twice, with a three-day gap in between. After collecting and analyzing the data, this study proved that Lex30 has a high degree of reliability. As Meara (2009, p. 46) states, "the correlation between the two sets of scores is .866 ( $p < .01$ )", which means that the "subjects taking the Lex30 test more than once at a given point in their L2 development will achieve broadly similar scores each time".

After proving that Lex30 is a reliable task that measures the productive vocabulary size, Fitzpatrick and Meara (2004) then proved that Lex30 is a valid task through two studies. The first study compared the performance of

46 adult L1 speakers of English with 46 non-native speakers. After collecting the data, two points were observed. The first was that the responses of the native speakers to the stimulus words differed from those of the non-native speakers in that the native speakers produced a high rate of low-frequency words. The second observed point was that 18 non-native speakers achieved a higher score than the native speakers. In addition, six native speakers reached a higher score compared with the highest scores of the non-native speakers, and this was due to the nature of the non-native speakers: they were teachers of English at an Icelandic secondary school.

The second study involved 55 Chinese learners of English, varying from intermediate to advanced. They translated a set of 60 Chinese words from their L1 to English, the target language, and they also completed the productive version of the VLT. When Fitzpatrick and Meara (2004) analyzed the data of this study, they found a high correlation between the tests; most importantly, there was a modest correlation with Lex30.

Although Lex30 has been validated, a reliability test was also run using SPSS. As Table 5 shows, the Cronbach's alpha is .930, which indicates a highly reliable test.

Table 5. Lex30 reliability

Cronbach's alpha	Cronbach's alpha based on standardized items	N. of items
.930	.694	5

### 2.3.2 COLLEX Test

One possible way to deal with the depth of vocabulary knowledge is by assessing learners' collocational knowledge (Gyllstad, 2007). As Milton (2009, p. 149) notes, "depth is generally used to refer to a wide variety of word characteristics, including the shades of meaning a word may carry, its connotations and collocations, the phrases and patterns of use it is likely to be found in, and the associations the word creates in the mind of the user." Moreover, most collocational studies in the literature concentrate on the content words (Gyllstad, 2005, 2007), whereas delexical verbs, which are verbs used with nouns as their objects to indicate simply that someone performs an action, are difficult to identify as collocations, even by advanced learners of English: i.e., *have, take, make, give, go, and do* (Altenberg & Granger, 2001; Nesselhauf, 2004).

Thus, Gyllstad (2007) developed a method to assess the depth of vocabulary knowledge by evaluating collocational knowledge. This method is known as the COLLEX test. The presented research adapted the fifth version of the COLLEX test to measure learners' depth of vocabulary knowledge (Appendix 3A contains the full version of the test). COLLEX 5 includes a set of 40 collocations, and each set has three options. Learners in this test have to select the most real collocation out of the three options, as shown below:

<i>do a visit</i>	<i>hit a visit</i>	<i>pay a visit</i>
-------------------	--------------------	--------------------

In addition, as Gyllstad (2007, p. 163) notes, two main points are specified with regard to COLLEX 5. The first point is that "only verb + NP items were used, which means that adjective + NP items were discarded" from the other versions. The second point is that "in terms of item-total correlation and item facility values, new items were created by adding a second distractor to each item", which means that the previous version of COLLEX 5 only had two options, and one of these should be selected, as shown below:

<i>solve a problem</i>	<i>break a problem</i>
------------------------	------------------------

In addition, the high-frequency words used in COLLEX 5 were chosen carefully through the JACET 8000, which is "a radically new word list designed for all English learners in Japan. This list is based on two kinds of corpora: the British National Corpus (BNC) and JACET 8000 sub-corpus" (Ishikawa, Uemura, Kaneda, Shimizu, Sugimori, & Tono, 2003). As Gyllstad (2007, p. 164) proposed, "a total of 112 different words (72 verbs and 40 nouns) were used, and 88 per cent of these words came from the 1-3K bands." Furthermore, as Gyllstad (2007, p. 164) argues, COLLEX 5 includes some lower-frequency words, and the reason behind this is "to make the distractors plausible" (Appendix 3B contains the word frequencies of the test).

In terms of reliability, literature has shown that all versions of the COLLEX text were reliable, as Table 6 demonstrates.

Table 6. COLLEX test reliability (Gyllstad, 2007, p. 154; Schmitt, 2010, p. 234)

Test version	Reliability
COLLEX 1	.54
COLLEX 2	.82
COLLEX 3	.84
COLLEX 4	.91
COLLEX 5	.89

### 2.3.3 Data Collection Procedures

The study took place in March during the second semester of the 2018–2019 academic year. The process of collecting the data began by asking the participants in the two groups to complete three tests. Confidentiality, privacy, and the right of withdrawal were explained to them. The first test was Meara and Jones's (1990) EVST, which was used to measure their receptive vocabulary. The second test was Meara and Fitzpatrick's (2000) Lex30 word association task, which was used to measure their productive vocabulary. The third test was Gyllstad's (2007) COLLEX test, which was used to investigate the depth of their vocabulary.

Of all three tests, only Lex30 had to be completed within a limited amount of time in order to evaluate the variation of responses between the intermediate and advanced groups. Therefore, participants were given 18 minutes to finish it. The EVST and the COLLEX test were not time limited; however, the participants were asked to finish within 30 minutes to avoid taking a long time.

### 2.3.4 Scoring and Statistical Tools Used

The criteria for scoring the three tests are totally different. First, as discussed before, Meara and Jones's (1990) EVST contains 300 words divided into five levels, with each level containing 40 real words and 20 pseudo-words. Therefore, this test has four possible marks, i.e., hit, false alarm, miss, and correct rejection. The maximum score for each block is out of 60 marks, as shown below:

- *Hit: ticking a real word = one mark*
- *False alarm: ticking a pseudo-word = no mark*
- *Miss: not ticking a real word = no mark*
- *Correct rejection: not ticking a pseudo-word = one mark*

Second, there are a number of steps to score the responses to Meara and Fitzpatrick's (2000) Lex30 word association task, which is a list of 30 stimulus words that require learners to produce a range of one to four responses or associations, as discussed before. The first step is to discard the stimulus words. For example, in the following example, the word *animal* should be discarded, and the words *elephant*, *tiger*, *farm*, and *wild* should be analyzed.

- *animal: elephant – tiger – farm – wild*

After discarding the stimulus words, the responses are typed into a machine-readable file. Responses with inflectional suffixes (plural forms, past tenses, past participle aspects, comparatives, superlatives, etc.) and frequent regular derivational affixes (-er, -less, -ness, -able, -ly, -ish, etc.) are lemmatized and counted as base forms of words, whereas responses with unusual affixes are treated as separate words, not lemmatized. At this stage, learners were given their scores out of 120.

After the responses are typed into a machine-readable file, they are analyzed by considering them as a generated text made by each learner. This is then typed into a computer program that analyzes texts according to the frequency level of each word. The current study classified words according to the largest available lists of word families, the British National Corpus/Corpus of Contemporary American English (BNC/COCA) word frequency lists (Nation, 2016), by using one of the most popular profiling programs known as Compleat Web VP V.2. The BNC/COCA word frequency lists "consist of 32-word family lists. Twenty-eight of the lists contain word families based on frequency and range data. The four additional lists are (1) an ever-growing list of proper names, (2) a list of marginal words including swear words, exclamations, and letters of the alphabet, (3) a list of transparent compounds, and (4) a list of acronyms" (Nation, 2016, p. 132).

When it comes to the mental lexicon, the word association responses are analyzed using Moreno Espinosa's (2009) adaptation of Fitzpatrick's (2006) model of word association. That is, test takers' responses to the stimulus words in Meara and Fitzpatrick's (2000) Lex30 word association task are classified into five different categories: syntagmatic responses, paradigmatic, clang, misunderstanding, and uninterpretable, as shown in



Table 7. Definition of word association categories (Moreno Espinosa, 2009, p. 101)

#	Category	Main features
1	Syntagmatic	<b># Position-based associations that include:</b> - consecutive <i>yx</i> , collocation ..... (e.g., teacher – school) - phrasal <i>xy</i> collocation ..... (e.g., dog – bark) - and different word class collocation ..... (e.g., sun – hot)
2	Paradigmatic	<b># Meaning-based associations that include:</b> - synonymy ..... (e.g., quick – fast) - antonymy ..... (e.g., silly – intelligent) - hyponymy ..... (e.g., cat – animal) - co-hyponymy ..... (e.g., table – chair) - hypernymy ..... (e.g., fruit – apple) - meronymy ..... (e.g., bedroom – house)
3	Clang	<b># Form-based associations including:</b> - orthographic ..... (e.g., there – three) - phonological associations ..... (e.g., dog – bog, think – thank) - responses that are simply a repetition of the stimulus word
4	Misunderstanding	<b># Responses based on:</b> - false cognates - mistaken cues for another English word
5	Uninterpretable	- No clear relationship has been identified between the prompt and the word association response.

Note. x = stimulus word, y = word association

Third, Gyllstad's (2007) COLLEX test (version 5) is a collocational test that contains a set of 40 collocations. Each set has three options, of which one is the correct combination of words that represents the real collocation. The scoring of this test is done by counting the correct choices of the collocations. The maximum mark in this test is 40.

The data were analyzed statistically by comparing the mean and calculating the percentages of both the intermediate and advanced groups using SPSS. The responses to the first and the third tests, the EVST and the COLLEX test, were analyzed using SPSS, whereas the responses to the second test, the Lex30 word association task, were analyzed using both a computer program called VocabProfile and SPSS.

The results for both the EVST and the Lex30 word association task were used to answer the first question of the study, and the result of the COLLEX test was used to answer the second question of the study. After this, a comparison was made between all the previous results to answer the third question of the study. Next, the responses to the Lex30 word association task were compared in order to answer the fourth and final question of the study.

### 3. Results and Discussion

The criterion used to assess the results of the three tests is the percentage score, as shown in Table 8.

Table 8. Description of the results

#	Score percentage	Description
1	1–19.9%	VERY LOW
2	20–39.9%	LOW
3	40–59.9%	MODERATE
4	60–79.9%	HIGH
5	80–100%	VERY HIGH

#### 3.1 Question One: What Is the Effect of General Language Proficiency on the Sizes of the Receptive and Productive Vocabularies of Learners of English at the University of Bahrain?

To answer this question, both the receptive and productive vocabularies of the intermediate and advanced groups were assessed using Meara and Jones's (1990) EVST and Meara and Fitzpatrick's (2000) Lex30 word association task. After using these two methods, the participants' responses provided in the Lex30 word association task (Meara & Fitzpatrick, 2000) were classified into different lexical categories using a computer program called Compleat Web VP V.2.

### 3.1.1 Eurocentres Vocabulary Size Test Results (Receptive V.)

Table 9 shows the findings of Meara and Jones's (1990) EVST for both the intermediate and advanced groups. It illustrates a number of items: the number of participants who took part in the test; the participants' minimum, maximum and mean scores; and standard deviations.

Table 9. EVST results of the intermediate and advanced groups

Group	N	Minimum	Maximum	Mean & percentage	SD
Intermediate	60	82	294	233 (77.67%)	46.89
Advanced	60	212	289	267 (89%)	14.91
<b>Difference in mean score and percentage</b>				34 (11.33%)	

Note. The highest possible score in the test is 300.

As can be seen in Table 9, the mean scores of the intermediate and advanced groups in Meara and Jones's (1990) EVST are slightly different: the overall level of the intermediate group is high (77.67% – M = 233), whereas the overall level of the advanced group is very high (89% – M = 367). The high score level of the advanced students is likely to be the effect of language proficiency. The standard deviation, however, indicates that the advanced-level students are more homogeneous as a group (SD = 14.91) than the intermediate-level students (SD = 46.89).

### 3.1.2 Word Association Task Results (Productive V.)

Table 10 shows the results of Meara and Fitzpatrick's (2000) Lex30 word association task for both the intermediate and advanced groups. It contains several items: the number of participants who took part in the test; the minimum and maximum numbers of responses provided by the participants; the participants' response means and standard deviations; and the difference between the means of the two groups.

Table 10. Lex30 results of the intermediate and advanced groups

Group	N	Minimum	Maximum	Mean & percentage	SD
Intermediate	60	18 (15%)	97 (80.83%)	63.56 (52.91%)	16.28
Advanced	60	43 (35.83%)	120 (100%)	98.60 (81.67%)	19.57
<b>Difference in mean score and percentage</b>				35.0333 (29.19%)	

Note. The highest possible score in the test is 120.

According to Table 10, the mean score of the productive vocabulary that test takers in the intermediate group provided in Meara and Fitzpatrick's (2000) Lex30 word association task was moderate (52.91%), whereas the mean score of the advanced group was very high (81.67%). In addition, an independent samples t-test was conducted to compare the numbers of responses by the two groups included in the analysis. The results in Table 1 and Table 11 show that there was a significant difference between the number of responses from the intermediate group (M = 63.56, SD = 16.28) and the advanced-level students (M = 98.6, SD = 19.57) (conditions:  $t(118) = 10.665, p = 0.000$ ).

Table 11. Independent samples test for number of responses

		Levene's test for equality of variances		t-test for equality of means			
		F	Sig.	t	df	Sig. (2-tailed)	Mean differenceStd. error difference
<b>N. of responses</b>	EVA	1.659	.200	10.665	118	.000	35.03333 3.28501
	EVNA			10.665	114.268	.000	35.03333 3.28501

Note. EVA: Equal variances assumed; EVNA: Equal variances not assumed.

### 3.1.3 Compleat VocabProfile V.2 Results (Productive V.)

The same words used in Meara and Fitzpatrick's (2000) Lex30 word association task were included in the Compleat VocabProfile V.2 to classify them into frequency categories based on the BNC/COCA (Nation, 2016) word frequency lists. The results of the Compleat VocabProfile V.2 are shown in Table 12 and Table 13. First,

Table 12 presents a comparison between the mean scores of the Tokens, Types, and Word Families provided by both the intermediate group and the advanced group in Meara and Fitzpatrick's (2000) Lex30 word association task.

Table 12. Mean scores of tokens, types, and word families

Groups	Tokens M.	Types M.	Word Families M.
Intermediate Group	62.05	59.72	57.62
Advanced Group	95.52	89.93	87.73
Difference	33.47 (53.94%)	30.22 (50.58%)	30.12 (52.25%)

The comparison in Table 12 shows that the mean score of the Tokens given by the intermediate group is 62.05, whereas the advanced group's mean score is 95.52. The mean score of the Types provided by the intermediate group is 59.72, and 89.93 for the advanced group. Moreover, the mean score of the Word Families elicited by the intermediate group is 57.62, whereas the mean score of the advanced group is 87.73. Thus, there was a significant difference between these figures for the intermediate group and the advanced group, as demonstrated in the following table.

Table 13. Independent samples test for number of tokens, types, and word families

		Levene's test for equality of variances		t-test for equality of means				
		F	Sig.	t	df	Sig. (2-tailed)	Mean difference	Std. error difference
<b>N. of Tokens</b>	EVA	1.818	.180	10.603	118	.000	33.46667	3.15637
	EVNA			10.603	114.496	.000	33.46667	3.15637
<b>N. of Types</b>	EVA	.461	.498	10.317	118	.000	30.21667	2.92872
	EVNA			10.317	117.066	.000	30.21667	2.92872
<b>N. of Word Families</b>	EVA	.444	.507	10.640	118	.000	30.11667	2.83052
	EVNA			10.640	116.914	.000	30.11667	2.83052

Note. EVA: Equal variances assumed; EVNA: Equal variances not assumed.

Table 13 indicates that there was a significant difference in the number of Tokens, Types, and Word Families for the intermediate and advanced groups (Tokens:  $t(118) = 10.603$ ,  $p = 0.000$ ; Types:  $t(118) = 10.317$ ,  $p = 0.000$ ; and Word Families:  $t(118) = 10.640$ ,  $p = 0.000$ ).

In addition, Table 14 demonstrates the distribution of the participants' words provided in Meara and Fitzpatrick's (2000) Lex30 word association task. These words were profiled according to the BNC/COCA (Nation, 2016) word frequency lists using Compleat VocabProfile V.2. The table compares the number of words produced by both the intermediate group and the advanced group in a classification of 25 categories (K1 to K25).

Table 14. Compleat Web VP V.2 classification

Category	Intermediate Group		Advanced Group		Status	
	Responses	Percentage	Responses	Percentage	Increase?	Percentage
<b>K1</b>	2,536	66.49%	3,449	58.3%	+ 913	36%
<b>K2</b>	707	18.54%	1,139	19.25%	+ 432	61.1%
<b>K3</b>	101	2.65%	372	6.29%	+ 271	268.31%
<b>K4</b>	133	3.49%	281	4.75%	+ 148	111.27%
<b>K5</b>	129	3.38%	235	3.97%	+ 106	82.17%
<b>K6</b>	67	1.67%	135	2.28%	+ 68	101.49%
<b>K7</b>	6	0.16%	24	0.41%	+ 18	300%
<b>K8</b>	14	0.37%	28	0.47%	+ 14	100%
<b>K9</b>	5	0.13%	16	0.27%	+ 11	220%
<b>K10</b>	8	0.21%	18	0.3%	+ 10	125%
<b>K11</b>	6	0.16%	11	0.19%	+ 5	83.33%
<b>K12</b>	4	0.1%	4	0.07%	=	0%
<b>K13</b>	3	0.8%	6	0.1%	+ 3	100%
<b>K14</b>	0	0%	0	0%	=	0%
<b>K15</b>	1	0.3%	2	0.03%	+ 1	100%
<b>K16</b>	1	0.3%	2	0.03%	+ 1	100%
<b>K17</b>	0	0%	2	0.03%	+ 2	200%
<b>K18</b>	0	0%	3	0.05%	+ 3	300%
<b>K19</b>	0	0%	0	0%	=	0%
<b>K20</b>	0	0%	1	0.02%	+ 1	100%
<b>K21</b>	1	0.3%	1	0.02%	=	0%
<b>K22</b>	0	0%	1	0.02%	+ 1	100%
<b>K23</b>	0	0%	0	0%	=	0%
<b>K24</b>	0	0%	1	0.02%	+ 1	100%
<b>K25</b>	0	0%	0	0.0%	=	0%
<b>Off-List</b>	92	2.41%	185	3.13%	+92	101.08%
<b>Total</b>	<b>3,814</b>		<b>5,916</b>		<b>+ 2,104</b>	<b>55.11%</b>

According to Table 14, the participants in the intermediate group produced 3,814 responses in the Lex30 word association task, whereas the participants in the advanced group produced 5,916 responses. The comparison between the responses of the two groups shown in the table indicates that there is a huge expansion of responses mainly in the first six categories of the BNC/COCA word frequency list: K1 increased by 36% (913 responses), K2 by 61.1% (432 responses), K3 by 268.31% (271 responses), K4 by 111.27% (148 responses), K5 by 82.17% (106 responses), and K6 by 101.49% (68 responses).

### 3.1.4 Discussion of the Results for Question One

Based on a comparison between the sizes of the receptive and productive vocabularies, several major points were shown. First, the use of the EVST revealed that learners of English at the University of Bahrain have a distinctive receptive vocabulary: the intermediate group's mean score for their receptive vocabulary was quite remarkable. The mean score of the participants in this group was high (233 = 77.67%), and the advanced group's mean score was very high (267 = 89%). These figures demonstrate that general language proficiency has a positive effect on the development of receptive vocabulary. Second, the use of the Lex30 word association task supports the fact that general language proficiency plays a significant role in developing the size of learners' vocabulary. When comparing the mean scores of the intermediate and advanced groups, the effect of general language proficiency becomes obvious: the mean score for their productive vocabulary increased from moderate (63.56 = 52.91%) to very high (98.6 = 81.67%). In addition, the classification of words in the Compleat VocabProfile V.2 indicates that this growth in vocabulary size was expanded among different categories of word frequency.

Overall, the intermediate and advanced groups' performance of vocabulary knowledge supports Meara's (2009) and Milton's (2009) assumption; that is, a learner's receptive vocabulary is larger than their productive vocabulary. This corroborates Ab Manan, Azizan, Fatima and Mohd's (2017) study: when they investigated the levels of receptive and productive vocabularies, they found that the former was between 2,000- and 3,000-word families, and the latter was around 2,000 word families. Likewise, Karakoç and Köse (2017) found that the size of their participants' receptive vocabulary (M = 87.18) was larger than the size of their productive vocabulary (M = 45.14).

### 3.2 Question Two: How Does General Language Proficiency Affect the Depth of Vocabulary Knowledge of English Learners at the University of Bahrain?

In order to answer question two, the depth of vocabulary variation of both the intermediate and advanced groups was investigated using Gyllstad's (2007) COLLEX test.

#### 3.2.1 COLLEX Test Results

Table 15 shows the findings of the COLLEX test for both the intermediate and advanced groups. It illustrates many items, including the number of participants; the minimum and maximum scores achieved by the participants; and the means and standard deviations of the participants' scores.

Table 15. COLLEX test results for the intermediate and advanced groups

Group	N	Minimum	Maximum	Mean & percentage	SD
Intermediate	60	12 (30%)	37 (92.5%)	21.56 (53.89%)	6.65
Advanced	60	14 (35%)	39 (97.5%)	28.08 (70.19%)	5.85
<b>Difference between mean score and percentage</b>				6.52 (16.30%)	

Note. The highest possible score in the test is 40.

Table 15 shows that the intermediate group's mean score was 21.56 (53.89%) with a standard deviation of 6.65, while the advanced group's mean score was 28.08 (70.19%) with a standard deviation of 5.85. The difference between the mean scores was very low (6.52 = 16.30%).

#### 3.2.2 Discussion of the Results for Question Two

Based on the results of the COLLEX test for both the intermediate and advanced groups, it is clear that there was a very low improvement in the learners' depth of vocabulary (difference in mean score: 6.52). The results of the COLLEX test show that the role of general language proficiency has developed the participants' depth of vocabulary from moderate to high: as Table 15 shows, the participants' mean score increased from 21.56 (53.89%) to 28.08 (70.19%).

With reference to the literature, Ebrahimi (2017) investigated depth of vocabulary using a number of tests, one of which was a productive test of collocation. The results of the test showed that learners' knowledge of collocations was considerably low. Furthermore, Bagherzadeh Hosseini and Akbarian (2007) evaluated the relationship between collocational competence and general language proficiency and found that these two variables are much related to each other.

### 3.3 Question Three: What Is the Relationship Between Receptive and Productive Vocabularies and the Depth of Vocabulary Knowledge?

To answer question three, the mean scores of the three methods used in questions one and two were compared to investigate the correlation between the receptive and productive vocabularies and the depth of vocabulary. In other words, answering question three required an investigation of the correlations between the intermediate and advanced groups' mean scores in the EVST, the Lex30 word association task, and the COLLEX test.

#### 3.3.1 Correlation Results

Table 16 illustrates the findings of the correlation test between the mean scores of the receptive vocabulary test (Meara and Jones's (1990) EVST), the productive vocabulary test (Meara and Fitzpatrick's (2000) word association task), and the depth of vocabulary test (Gyllstad's (2007) COLLEX test).

Table 16. Results of the Pearson correlation between the tests' variable mean scores

Tests		Lex30 task	EVST	COLLEX test
EVST	P. correlation	.984	1	.947
	Sig. (2-tailed)	.115		.209
	N	3	3	3
Lex30 task	P. correlation	1	.984	.989
	Sig. (2-tailed)		.115	.094
	N	3	3	3
COLLEX test	P. correlation	.989	.947	1
	Sig. (2-tailed)	.094	.209	
	N	3	3	3

To illustrate, starting with the first method used in the study, the correlation between the EVST and the Lex30 word association task and the COLLEX test was 0.984 and 0.947, respectively. In addition, the table shows that the correlation between the Lex30 word association task and the EVST and the COLLEX test was 0.984 and 0.989, respectively. In addition, the correlation between the COLLEX test and the Lex30 word association task and the EVST was 0.989 and 0.947, respectively.

The results indicate that the tests' three mean scores are positively correlated. In other words, the table demonstrates that the sizes of the receptive and productive vocabularies and the depth of vocabulary are significantly correlated.

### 3.3.2 Discussion of the Results for Question Three

On the basis of the correlation results shown in Table 16, question three can be answered logically by stating that, as there are significant correlations between the results of the three tests, the relationship between the sizes of the receptive and productive vocabularies and the depth of vocabulary knowledge is positively correlated. This finding is consistent with a number of studies. For example, Shin, Chon and Kim (2011) demonstrated that, overall, their participants' receptive and productive vocabulary sizes indicated moderate correlations when an investigation was undertaken to provide an assessment of their learners' vocabulary sizes. In addition, a similar finding was also reported in Fleckenstein (2018): a strong, positive relationship was revealed when these two aspects of vocabulary knowledge were assessed. Also, on the theme of size and depth of vocabulary knowledge, Bardakçı (2016) found that these two dimensions are closely correlated, i.e., these dimensions have significant effects on learners' vocabulary profiles.

### 3.4 Question Four: What Is the Relationship Between Vocabulary Size and the Nature of Lexical Networking?

In order to answer question four, the Lex30 word association task responses of the intermediate and advanced groups were classified according to Fitzpatrick's (2006) model of analyzing word association categories and sub-categories. As discussed before, five classifications were used in this model: syntagmatic, paradigmatic, clang, misunderstanding, and uninterpretable associations.

#### 3.4.1 Classification of the Word Associations

Table 17 shows the results of classifying the intermediate and advanced groups' responses to the Lex30 word association task. It illustrates many items, including the size of the productive vocabulary and the classification of the responses in accordance with Moreno Espinosa's (2009) adaptation of Fitzpatrick's (2006) model, which makes the explanation of the classification clearer.

Table 17. Size and classification of the word association task responses

Groups	N	Size	S	P	C	M	U
Intermediate	60	3,814	1,448 (37.97%) M: 24.13	2,023 (53.04%) M: 33.71	43 (1.13%) M: 1.43	116 (3.04%) M: 3.13	184 (4.82%) M: 4
Advanced	60	5,916	1,922 (32.49%) M: 32.03	3,446 (58.25%) M: 57.43	48 (0.81%) M: 1.54	108 (1.83%) M: 2.35	392 (6.63%) M: 7

Note. S: Syntagmatic, P: Paradigmatic, C: Clang, M: Misunderstanding, and U: Uninterpretable.

On the one hand, as seen in Table 17, the intermediate group provided 3,814 responses in the Lex30 word association task. These responses were analyzed and classified as follows: 53.04% of the responses were paradigmatic associations, 37.97% were syntagmatic associations, 4.82% were uninterpretable associations, 3.04% were misunderstanding associations, and 1.13% were clang associations. On the other hand, the advanced group provided 5,916 responses: 58.25% were paradigmatic associations, 32.49% were syntagmatic associations, 6.63% were uninterpretable associations, 1.83% were misunderstanding associations, and 0.81% were clang associations.

#### 3.4.2 Discussion of the Results for Question Four

According to Table 17, the responses of the intermediate and advanced groups were similar to each other. Most responses were paradigmatic, then syntagmatic, and the least popular responses were clang; no shifts were identified among all the responses. What is interesting here is that, although the effect of general language proficiency affected the sizes of the responses, the classification of the responses did not change among the

categories of Moreno Espinosa's (2009) adaptation of Fitzpatrick's (2006) word association model. Therefore, question four can be answered by saying that there is no relationship between the size of vocabulary and the nature of lexical networking. However, the figures shown in Table 17 support Meara's (2009, p. 6) assumption that "normal adults produce two main types of association", syntagmatic and paradigmatic, and they "produce more paradigmatic responses than syntagmatic ones." Moreover, these figures are interestingly in line with Moreno Espinosa's (2009) findings when the researcher investigated learners' lexical profiles using the Lex30 word association task (Meara & Fitzpatrick, 2000): the vast majority of responses were syntagmatic, then paradigmatic, and the least were clang.

#### 4. Conclusion

This study attempted to answer four research questions related to the size and depth of vocabulary knowledge. The first and second research questions were aimed at evaluating the effect of general language proficiency on the development of the vocabulary size and depth of knowledge of EFL learners at the University of Bahrain. The third research question was aimed at investigating the relationship between vocabulary size and depth of knowledge. Finally, the fourth research question was aimed at investigating the relationship between the size of vocabulary and the organization of lexical networking in relation to general language proficiency.

The results indicate that general language proficiency has a prime and positive effect on the size and depth dimensions of vocabulary knowledge. According to the findings, the size of the learners' receptive vocabulary changed from high (77.67%) to very high (89%), and the size of their productive vocabulary changed from moderate (52.91%) to very high (81.67%). In addition, the learners' depth of vocabulary changed from moderate (53.89%) to high (70.19%). The study also shows that there is a positive correlation between vocabulary size and depth of knowledge. However, there is no evident relationship between the size of vocabulary and the nature of lexical networking; the order of responses by the participants in the Lex30 word association task did not change or shift.

##### 4.1 Limitations

Using hardcopies of the tests was a major limitation in the presented study, with the researcher transferring all the participants' responses into software programs in order to find the answers to the research questions. Unfortunately, the computerized version of the Lex30 word association task used in the study to assess learners' productive vocabulary size was, for various reasons, no longer supported by the owner. Therefore, the time required for analyzing the findings was extended.

##### 4.2 Recommendations

Based on the investigation into vocabulary size and depth of knowledge in the current study, several recommendations can be made to help develop EFL learners' L2 vocabulary learning competence.

The receptive vocabulary sizes of the intermediate and advanced groups were significantly distinctive: the intermediate group's mean score was high (77.67%), and the advanced group's mean score was very high (89%). It seems worthwhile to recommend that qualitative studies are carried out to investigate and then generalize the strategies for learning receptive vocabulary used by EFL learners at the University of Bahrain. In terms of their productive vocabulary size, the intermediate and advanced groups' mean scores demonstrate the need to focus on teaching more productive vocabulary: the intermediate group's mean score was moderate (52.91%), and the advanced group's mean score was very high (81.67%). It would also be worthwhile to investigate the mechanism that stimulates EFL learners to use their receptive vocabulary in writing and speaking. Furthermore, the difference between the two groups' mean scores for vocabulary depth reveals that EFL learners should be exposed to learning more semantic relationships, i.e., word associations and collocations.

Further research could also be conducted to assess the exact size of the receptive vocabulary knowledge of EFL learners at the University of Bahrain using Nation and Beglar's (2007) VST. Once vocabulary knowledge has been assessed, a valuable overview of learners' vocabulary competence will be revealed to the instructors and curriculum designers engaged in building and developing English courses. It could be argued that English learning courses would be more relevant if an assessment of vocabulary was included at the beginning and end of these courses. Whether these courses cover listening, reading, writing, or speaking skills, an overview of the vocabulary assessment has considerable benefits. In addition, the assessment of EFL learners' vocabulary knowledge could also be beneficial when compared with that of other learners around the world. The comparison explores major points on the strengths and weaknesses of the vocabulary learning process. Furthermore, it would be beneficial to investigate vocabulary knowledge in relation to different gender perspectives alongside the language proficiency levels of EFL learners in Bahrain.

More broadly, greater efforts are needed to build a vocabulary assessment application for the University of Bahrain to be used with the admission aptitude test and personal interview.

## References

- Ab Manan, N. A., Azizan, N., Fatima, N., & Mohd, W. (2017). Receptive and productive vocabulary level of diploma students from a public university in Malaysia. *Journal of Applied Environmental and Biological Science*, 7(1S), 53–59.
- Alavi, S., & Akbarian, I. (2012). The role of vocabulary size in predicting performance on TOEFL reading item types. *System*, 40(3), 376–385. <https://doi.org/10.1016/j.system.2012.07.002>
- Altenberg, B., & Granger, S. (2001). The grammatical and lexical patterning of MAKE in native and non-native student writing. *Applied Linguistics*, 22(2), 173–195. <https://doi.org/10.1093/applin/22.2.173>
- Anderson, R. C., & Freebody, P. (1983). Reading comprehension and the assessment and acquisition of word knowledge. *Advances in Reading/Language Research*, 2, 231–256.
- Baba, K. (2002). Test review: Lex 30. *Language Testing. Update*, 32, 68–71.
- Bagherzadeh Hosseini, M. S., & Akbarian, I. (2007). Language proficiency and collocational competence. *The Journal of Asia TEFL*, 4(4), 35–58.
- Bardakçi, M. (2016). Breadth and depth of vocabulary knowledge and their effects on L2 vocabulary profiles. *English Language Teaching*, 9(4), 239. <https://doi.org/10.5539/elt.v9n4p239>
- Çakmak, F., Namaziandost, E., & Kumar, T. (2021). CALL-Enhanced L2 Vocabulary Learning: Using Spaced Exposure through CALL to Enhance L2 Vocabulary Retention. *Education Research International*, 1–8. <https://doi.org/10.1155/2021/5848525>
- Ebrahimi, A. (2017). *Measuring productive depth of vocabulary knowledge of the most frequent words*. Master's Degree. University of Western Ontario.
- Fitzpatrick, T. (2006). Habits and rabbits: Word associations and the L2 lexicon. In S. H. Foster-Cohen, M. Medved Krajnović & J. Mihaljević Djigunović (Eds.), *EUROSLA yearbook* (Vol. 6, pp. 121–145). Amsterdam: John Benjamins Publishing Company. <https://doi.org/10.1075/eurosla.6.09fit>
- Fitzpatrick, T., & Meara, P. M. (2004). Exploring the validity of a test of productive vocabulary. *Vigo International Journal of Applied Linguistics*, 1, 55–74.
- Fleckenstein, L. (2018). *English vocabulary knowledge at the end of compulsory education in Iceland: Exploring receptive and productive English vocabulary in relation to extramural activities among 10th Grade students in Iceland*. Retrieved from <https://skemman.is/bitstream/1946/30002/1/Meistararitgerð%20Fleckenstein.pdf>
- Gyllstad, H. (2005). *Devising and evaluating tests of collocation knowledge: COLLEX and COLLMATCH*. Paper presented at VARG Summer School 2005, Swansea, Wales, UK.
- Gyllstad, H. (2007). *Testing English collocations: Developing receptive tests for use with advanced Swedish learners*. Språk-och litteraturcentrum, Lunds universitet.
- Ha, H. (2021). Exploring the relationships between various dimensions of receptive vocabulary knowledge and L2 listening and reading comprehension. *Language Testing in Asia*, 11(1). <https://doi.org/10.1186/s40468-021-00131-8>
- Ishikawa, S., Uemura, T., Kaneda, M., Shimizu, S., Sugimori, N., & Tono, Y. (2003). *JACET 8000: JACET list of 8000 basic words*. Tokyo: JACET.
- Janebi Enayat, M., & Derakhshan, A. (2021). Vocabulary size and depth as predictors of second language speaking ability. *System*, 99. <https://doi.org/10.1016/j.system.2021.102521>
- Kalay, D (2021). Receptive Vocabulary Knowledge of University Students in CLIL Instruction. *European Journal of Foreign Language Teaching*, 5(3). <https://doi.org/10.46827/ejfl.v5i3.3527>
- Karakoç, D., & Köse, G. D. (2017). The impact of vocabulary knowledge on reading, writing and proficiency scores of EFL learners. *Journal of Language and Linguistic Studies*, 13(1), 352–378.
- Knight, S. (1994). Dictionary use while reading: The effects on comprehension and vocabulary acquisition for students of different verbal abilities. *The Modern Language Journal*, 78, 285–299. <https://doi.org/10.1111/j.1540-4781.1994.tb02043.x>
- Laufer, B. (1997). The lexical plight in second language reading: Words you don't know, words you think you



- know, and words you can't guess. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition: A rationale for pedagogy*. New York: Cambridge University Press. <https://doi.org/10.1017/CBO9781139524643.004>
- Laufer, B., & Nation, I. S. P. (1995). Lexical richness in L2 written production: Can it be measured? *Applied Linguistics*, 16(3), 307–322. <https://doi.org/10.1093/applin/16.3.307>
- Luo, Y., Song, H., Wan, L. & Zhang, X. (2021) The Effect of Vocabulary Depth and Breadth on English Listening Comprehension Can Depend on How Comprehension Is Measured. *Frontiers in Psychology*, 12. <https://doi.org/10.3389/fpsyg.2021.657573>
- Meara, P. M. (1990). Some notes on the Eurocentres vocabulary tests. In J. Tommola (Ed.), *Foreign language comprehension and production* (pp. 103–113). Turku: AFinLa Yearbook.
- Meara, P. M. (1996). The dimensions of lexical competence. In G. Brown, K. Malmkjaer & J. Williams (Eds.), *Performance and competence in second language acquisition* (pp. 35–53). Cambridge: Cambridge University Press.
- Meara, P. M. (2009). *Connected words: Word associations and second language vocabulary acquisition*. Amsterdam: John Benjamins. <https://doi.org/10.1075/llt.24>
- Meara, P. M., & Buxton, B. (1987). An alternative to multiple choice vocabulary tests. *Language Testing*, 4(2), 142–154. <https://doi.org/10.1177/026553228700400202>
- Meara, P. M., & Fitzpatrick, T. (2000). Lex30: An improved method of assessing productive vocabulary in an L2. *System*, 28(1), 19–30. [https://doi.org/10.1016/S0346-251X\(99\)00058-5](https://doi.org/10.1016/S0346-251X(99)00058-5)
- Meara, P. M., & Jones, G. (1988). Vocabulary size as a placement indicator. In P. Grunwell (Ed.), *Applied Linguistics in Society* (pp. 80–87). London: CILT.
- Meara, P. M., & Jones, G. (1990). *The Eurocentres Vocabulary Size Test*. 10KA. Zurich: Eurocentres.
- Milton, J. L. (2009). *Measuring second language vocabulary acquisition*. Clevedon, UK: Multilingual Matters. <https://doi.org/10.21832/9781847692092>
- Moreno Espinosa, S. (2009). Young learners' L2 word association responses in two different learning contexts. In Y. Ruiz de Zarobe & R. M. Jiménez Catalán (Eds.), *Content and language integrated learning: Evidence from research in Europe* (pp. 93–111). Clevedon: Multilingual Matters. <https://doi.org/10.21832/9781847691675-009>
- Nation, I. S. P. (1984). *Vocabulary lists: words, affixes, and stems*. Wellington: English University of Wellington, English Language Institute.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139524759>
- Nation, I. S. P. (2016). *Making and using word lists for language learning and testing*. Amsterdam: John Benjamins. <https://doi.org/10.1075/z.208>
- Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, 31(7), 9–13.
- Nation, I. S. P., & Webb, S. (2011). *Researching and analysing vocabulary*. Boston, MA: Heinle Cengage Learning.
- Nesselhauf, N. (2004). What are collocations? In D. J. Allerton, N. Nesselhauf & P. Skandera (Eds.), *Phraseological units: Basic concepts and their application* (pp. 1–21). Basel: Schwabe.
- Read, J. (2004). Plumbing the depths: How should the construct of vocabulary knowledge be defined. In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language* (pp. 209–227). Amsterdam: John Benjamins. <https://doi.org/10.1075/llt.10.15rea>
- Richards, J. C. (1976). The role of vocabulary teaching. *TESOL Quarterly*, 10(1), 77–89. <https://doi.org/10.2307/3585941>
- Richards, J. C., & Rodgers, T. S. (2001). *Approaches and methods in language teaching* (2nd ed.). Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511667305>
- Rixon, S. (2013). *British Council survey of policy and practice in primary English language teaching worldwide*. London: British Council.
- Schmitt, N. (2000). *Vocabulary in language teaching*. New York: Cambridge University Press.

- Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language Teaching Research*, 12(3), 329–363. <https://doi.org/10.1177/1362168808089921>
- Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Basingstoke, England: Palgrave Macmillan. <https://doi.org/10.1057/9780230293977>
- Schmitt, N. (2014). Size and depth of vocabulary knowledge: What the research shows. *Language Learning*, 64(4), 913–951. <https://doi.org/10.1111/lang.12077>
- Shin, D., Chon, Y. V., & Kim, H. (2011). Receptive and productive vocabulary sizes of high school learners: What next for the basic word list? *English Teaching*, 66(3), 123–148. <https://doi.org/10.15858/engtea.66.3.201109.123>
- Sims, V. M. (1929). The reliability and validity of four types of vocabulary test. *Journal of Educational Research*, 20(2), 91–96. <https://doi.org/10.1080/00220671.1929.10879971>
- Thorndike, E. L., & Lorge, I. (1944). *The teacher's word book of 30,000 words*. New York: Teacher's College, Columbia University.
- Tilley, H. C. (1936). A technique for determining the relative difficulty of word meanings among elementary school children. *Journal of Experimental Education*, 5(1), 61–64. <https://doi.org/10.1080/00220973.1936.11010031>
- Wero, Y., Machmud, K. & Husain, N. (2021) The Study on Students' Vocabulary Size. *Jambura Journal of English Teaching and Literature*, 2(1), 22–34. <https://doi.org/10.37905/jetl.v2i1.10279>
- Wilkins, D. A. (1972). *Linguistics in language teaching*. London: Edward Arnold.
- Yoshii, M., & Flaitz, J. (2013). Second language incidental vocabulary retention: The effect of text and picture annotation types. *CALICO Journal*, 20(1), 33–58. <https://doi.org/10.1558/cj.v20i1.33-58>
- Zimmerman, J., Broder, P. K., Shaughnessy, J. J., & Underwood, B. J. (1977). A recognition test of vocabulary using signal-detection measures and some correlates of word and non-word recognition. *Intelligence*, 1(1), 5–13. [https://doi.org/10.1016/0160-2896\(77\)90025-3](https://doi.org/10.1016/0160-2896(77)90025-3)

### Copyrights

Copyright for this article is retained by the author, with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).