The Relationship between Vocabulary Size and Depth for Iranian EFL Learners at Different Language Proficiency Levels

Mostafa Janebi Enayat [a], Seyed Mohammad Reza Amirian [b, *]

[a] *University of Maragheh, Iran*

[b] *Hakim Sabzevari University, Iran*

A B S T R A C T

The present study investigated the association between vocabulary size and depth of Iranian EFL learners at different language proficiency levels. Additionally, the extent that this relationship could be different for low- and high-word-frequency bands was probed. In so doing, the Word Associates Test (WAT), the Vocabulary Levels Test (VLT), the monolingual Vocabulary Size Test (VST) and its bilingual Persian version were administered to 122 Iranian English as a Foreign Language (EFL) learners (F=79, M=43) who had been classified into three language proficiency levels by means of administering the Oxford Quick Placement Test (OQPT). The findings indicated the following: (a) vocabulary size and depth had a significant correlation for the lower-intermediate students; (b) size and depth dimensions had a moderate association for the upper-intermediate participants; (c) vocabulary size and depth were not significantly correlated for the advanced EFL learners; (d) the relationship between the higher-frequency bands of vocabulary size and vocabulary depth was significant for lower proficiency levels; and (e) the lower-frequency vocabulary size was not correlated with vocabulary depth for any of the proficiency levels. The results have implications for vocabulary depth instruction for higher language proficiency levels and suggest teachers, test developers, and materials designers to incorporate the dimension of word associations into the construct of word knowledge.

Keywords: language proficiency; vocabulary knowledge; vocabulary size; vocabulary depth; EFL learners

* Corresponding author: Department of English Language & Literature, Hakim Sabzevari University, Sabzevar, Iran
Email address: smr.amirian@gmail.com

Introduction

Vocabulary knowledge has been considered as an important factor in language proficiency (Alavi & Akbarian, 2012; Amirian, Mallahi, & Zaghi, 2015) and lexical availability has been identified as a key factor in successful L2 production (Avila Muñoz, 2017). The reason behind this significance is the fundamental role of word knowledge in improving the proficiency in language skills (Zhang & Li, 2011). As argued by Schmitt (2010), "one thing that all of the partners involved in learning process (students, teachers, materials writers, and researchers) can all agree upon is that learning vocabulary is an essential part of mastering a second language" (p. 4).

The construct of lexical knowledge is not simple and "the exact nature of lexical knowledge has always perplexed researchers and teachers" (Schmitt, 2014, p. 914). Different frameworks have been proposed by scholars of the field of vocabulary knowledge to classify this construct (Henriksen, 1999; Nation, 2001). In one framework, knowledge of vocabulary has been categorized into size and depth (Haastrup & Henriksen, 2000; Read, 1993, 2000). Size of lexical knowledge refers to the quantity of words language learners know (Nation, 2001) and does not include word associations. Vocabulary depth, however, is about the quality of vocabulary knowledge which includes the associations among the lexical items as well (Read, 2000).

The fundamental role of word knowledge has been, therefore, a motive for its investigation in both the receptive and productive skills. The studies targeting the receptive skills were primarily focused on the reading skill (e.g., Alavi & Akbarian, 2012; Cheng & Matthews, 2018; Kaivanpanah & Zandi, 2009; Noro, 2002; Qian & Schedl, 2004; Shiotsu & Weir, 2007) with very few studies probing the role of the size and depth dimensions in L2 listening (e.g., Cheng & Matthews, 2015; Dabbagh, 2016; Matthews, 2018; Stæhr, 2008). These studies mostly reported that vocabulary size was a stronger predictor of L2 reading and listening than the depth of word knowledge. In comparison with the bulk of research on vocabulary in receptive skills, the ones focusing on productive skills have been scarce in the literature. Similarly, the findings of studies investigating the association between both size and depth aspects of vocabulary and L2 writing (e.g., Dabbagh & Janebi Enayat, 2019) and speaking (e.g., Alharthi, 2020) revealed that the size dimension had more shared variance with the dependent variable than the depth aspect.

The studies in the literature of L2 vocabulary size and depth have mostly examined the simultaneous contribution of these dimensions in L2 skills since they are highly interconnected (Ishii & Schmitt, 2009; Meara & Wolter, 2004; Schmitt, 2014). Despite the indisputable link between these aspects, the nature of their association across language proficiency levels has remained obscure. While some studies (e.g., Akbarian, 2010; Nurweni & Read, 1999) reported the association to be stronger for higher proficiency students, other researchers (e.g., Henriksen, 2008; Noro, 2002) demonstrated that the link got weaker as the learners improved in language proficiency. This study aimed at inspecting this interaction further by using three measures of vocabulary size to obtain more conclusive results. The nature of this interconnection was also scrutinized by comparing the correlations for high- and low-word-frequency bands.

Literature review

Vocabulary size and depth in second language skills

The unique characteristics of size and depth of word knowledge have been a motivation for many vocabulary researchers to investigate their contribution to second language (L2) skills. However, receptive skills, especially reading, has gained more attention in the related literature. The

relationship between vocabulary size and reading has been investigated in a number of studies. Results of these studies confirmed the significant positive relationship between the two constructs. Shiotsu and Weir (2007), for instance, investigated the extent that vocabulary size and grammatical knowledge could contribute to L2 reading. Using Structural Equation Modeling (SEM) as a better measure for finding the predictive ability of grammar and vocabulary size, the authors reported that grammatical knowledge predicted L2 reading over and above word knowledge.

The role of vocabulary size in L2 reading was examined by Zhang and Anual (2008) who obtained results showing a great explanatory power for vocabulary size in reading comprehension. They used Vocabulary Levels Test (VLT) to measure the participants' size of word knowledge and investigate how this aspect of vocabulary could affect secondary students' reading comprehension. The results of their research indicated that the students' knowledge of the 2000- and the 3000-word levels could predict their reading comprehension. Stæhr (2008) also inspected the correlation between EFL vocabulary size and reading comprehension in Denmark. This study found evidence for the significant relationship between the two variables.

Alavi and Akbarian (2012) further examined the predictive ability of L2 vocabulary size in performance on TOEFL reading items. Results indicated that the students' ability to answer reading comprehension items that checked comprehension of main idea and stated detail as well as the lexical inferencing ability was correlated with their scores on the VLT. For the participants at high vocabulary size, the only test item that correlated with the VLT was the guessing vocabulary item.

Nassaji (2006) investigated the extent that vocabulary depth could explain the types of lexical inferencing strategies used by ESL learners. Using Word Associates Test (WAT) to assess vocabulary depth, and introspective think-aloud technique to elicit inferencing strategies used by the students while reading, he found that the learners who possessed a higher depth of vocabulary knowledge were better users of lexical inferencing strategies.

In another study, Qian and Schedl (2004) investigated whether vocabulary depth could predict scores of TOEFL reading comprehension test and concluded that "in terms of item difficulty level and power of predicting reading performance – the DVK [Depth of Vocabulary Knowledge] format is worth further evaluation as a basis for developing potential item types in the context of TOEFL 2000" (p. 46). Contributing to the previous studies in this area, Kaivanpanah and Zandi (2009) concluded that language proficiency could affect the learners' performance on depth of vocabulary knowledge test in that it may "play a less conspicuous role in reading comprehension at lower proficiency levels because the depth of vocabulary knowledge is not well developed in such levels" (p. 700). In addition, their results revealed that grammatical knowledge caused greater variance in the participants' scores on reading than vocabulary depth.

In some other studies, the simultaneous examination of breadth and depth of vocabulary gained attention. Qian (2002) evaluated the contribution of these two dimensions of word knowledge to TOEFL reading comprehension. Using Pearson correlation analysis, his study showed that the two aspects of size and depth were interconnected and their combination resulted in a stronger prediction of L2 reading.

As for the role of size in L2 receptive skills, Stæhr (2008) found weaker correlation between test scores of vocabulary size and listening ($r = .69$) in comparison with the one between vocabulary size and reading ($r = .83$). In another study, Stæhr (2009) found that vocabulary size and depth

had significant correlations with listening ($r = .70$ and $r = .65$, respectively) and together could predict about 50% of the variance observed in the listening comprehension scores. Matthews and Cheng (2015) studied the extent that high- and low-frequency vocabulary could predict L2 listening. The researchers used IELTS listening tests and a dictation test designed to measure the high-frequency vocabulary and reported that knowledge of the 3,000-word-frequency level could explain more variance in L2 listening than the other two levels, suggesting that more vocabulary size led to a better listening comprehension.

Batty (2007) probed the role of vocabulary depth in L2 speaking and writing using the Kanda English Proficiency Test (KEPT) and reported that depth predicted the writing performance but not the scores in the oral section. Stæhr (2008) also examined the extent that writing and vocabulary size were associated with each other. Results of this study showed that L2 writing was related to the receptive vocabulary size of the participants, indicating the significant role of size of vocabulary in productive skills as well. In contrast, Baba (2009) studied the link between size and depth dimensions and EFL summary writing performance and reported an insignificant association between L2 summary writing and aspects of lexical knowledge. The author inferred that it "makes sense for such complicated tasks as summary writing, which entails a wider variety of cognitive skills than simply writing or reading" (p. 202).

The link between vocabulary size and depth

Using the VLT and WAT as measures of vocabulary size and depth, respectively, the association between these two dimensions have been approved in the literature (Schmitt, 2014). Qian (1999), for example, examined the contribution of vocabulary size and depth to L2 reading and reported a high correlation between these dimensions of word knowledge. Using the same tests of vocabulary size and depth, Gyllstad (2007) and Zhang (2012) also found high correlations between the two dimensions.

However, while the correlation between these two dimensions have been found to be stronger for higher proficiency levels by some studies, some others have demonstrated that this relationship is stronger for lower proficiency levels and weaker for students with larger vocabulary sizes. Nurweni and Read (1999) examined the L2 word knowledge of a group of EFL students by administering a translation task to measure their size and a word association test to test their depth of vocabulary knowledge. The results of their analysis showed that the relationship between the vocabulary size and depth was higher for the participants with higher proficiency and larger vocabulary sizes. Similarly, Akbarian (2010) examined the extent that vocabulary size and depth were linked for the Iranian ESP/EAP learners, and, using regression analyses, found that these dimensions of word knowledge had 46% shared variance for the low-proficient and 80% for the high-proficient students. In other words, the relationship was stronger for the students with higher language proficiency.

Contrary to the findings of the aforementioned studies, some others have reported that the link between the dimensions of size and depth is weaker for the higher proficient students. Shimamoto (2000), for instance, aimed at finding the extent that the VLT scores and a series of vocabulary depth tests tapping different types of word associations were correlated and reported that the relationship was weaker for the participants with a larger vocabulary size. In a similar study, Noro (2002) scrutinized the link between these two dimensions of word knowledge using the VLT and WAT and reported that, at lower proficiency levels, the size and depth aspects of vocabulary knowledge were more related. Similar results were found by Henriksen (2008) who examined the same relationship and compared the younger graders with the advanced-level students.

Overall, the above review shows that the relationship between the two aspects of size and depth of word knowledge is ambiguous when the variable of language proficiency intervenes. In fact, as pointed out by Schmitt (2014, p. 940), "at the moment, it is difficult to interpret this conflicting evidence and come to conclusions about the strength of the relationship between size and organization as vocabulary size grows". Taking into account this lack of agreement in the literature on the link between vocabulary size and depth across students at different language proficiency and/or vocabulary size, the current study made an effort to inspect the association between the two dimensions of vocabulary knowledge using a range of measures of vocabulary size to gain more reliable results. The relationship between the two aspects was further investigated for high- and low-word-frequency levels of the vocabulary size tests to find if the nature of this relationship varies as a matter of word-frequency band. These research questions were thus addressed in the present study:

1. Are vocabulary size and depth of Iranian EFL learners at different language proficiency levels correlated?

2. Are higher-frequency vocabulary size and depth of Iranian EFL learners correlated at different language proficiency levels?

3. Are lower-frequency vocabulary size and depth of Iranian EFL learners correlated at different language proficiency levels?

Method

Participants

The sample of this study included 122 undergraduate students majoring in Teaching English as a Foreign Language (TEFL) and English Literature at three state universities in Iran. The participants were 43 males and 79 females, and their average age was 22. Oxford Quick Placement Test (OQPT, 2004) was utilized to classify the test takers based on their language proficiency levels. Accordingly, they were grouped into lower-intermediate (31 females and 12 males), upper-intermediate (35 females and 22 males), and advanced (13 females and 9 males) students.

Instruments

In this research, the OQPT was used to identify the language proficiency level of the students and put them into three groups. To measure their vocabulary size, three tests of VST, bilingual Persian VST, and VLT were administered. The purpose of administering three tests of vocabulary size was to obtain a more reliable estimate of the participants' size or quantity of vocabulary knowledge. The WAT was also employed to assess the vocabulary depth of the students.

Oxford Quick Placement Test (OQPT, 2004)

The OQPT was used to check the homogeneity of the participants in terms of language proficiency and classify them into three levels. This test employs item formats like cloze and multiple-choice items to measure grammatical and lexical knowledge as well as the test takers' reading comprehension. OQPT has 60 items and has been identified as a reliable and valid placement test (Geranpayeh, 2003).

Vocabulary Levels Test (VLT)

The VLT was first developed by Nation (1983), but Schmitt, Schmitt, and Clapham (2001) modified this test by making the definitions shorter to increase the validity. Also, they compared the test to a parallel version to check the reliability of this measure of vocabulary size. It tests the learners' vocabulary size at four word-frequency bands. The VLT uses a multiple matching format that requires the test takers to match a cluster of definitions with some target words (see Figure 1). The test takers are advised not to do blind guessing, but informed guessing is encouraged. The reliability of this test has reached .92 in a study by Qian (1999).

    1.   accident

    2.   debt   …. loud deep sound

    3.   fortune   …. something you must pay

    4.   pride   …. having a high opinion of yourself

    5.   roar

    6.   thread

Figure 1. A VLT sample item.

Monolingual Vocabulary Size Test (VST)

Nation and Beglar (2007) designed the VST "to provide a reliable, accurate, and comprehensive measure of second language English learners' written receptive vocabulary size from the first 1000- to the fourteenth 1000-word families of English" (Beglar, 2010, p. 3). The word family lists of the British National Corpus (BNC) were used for the development of this test (Nation, 2006). It has 10 items at each word-frequency level, so, overall, it has 140 items for 14 word-frequency bands. A sample item is provided in Figure 2. Using Rasch analyses, Beglar (2010) validated the VST and reported that this test enjoyed high validity and reliability.

ALUM: This contains **alum**.

    a.   A poisonous substance from a common plant

    b.   A soft material made of artificial threads

    c.   A tobacco powder once put in the nose

    d.   A chemical compound usually involving aluminium

Figure 2. An example item from the VST.

Bilingual Persian VST

Karami (2012) designed and validated the bilingual Persian version of the VST. Using factor analysis and one-way between subjects ANOVA, the researcher found that the test enjoyed a high

level of reliability and validity and distinguished students at different language proficiency levels. Figure 3 demonstrates a sample item from this version of the VST.

ALUM: This contains **alum**.

a.   ماده سمی که از گیاهی معروف به دست می‌آید

b.   پارچه‌ای که از الیاف مصنوعی درست شده است

c.   نوعی ماده‌ی مخدر

d.   نوعی ترکیب شیمیایی

Figure 3. An example item from the bilingual Persian version of VST.

Word Associates Test (WAT)

Developed by Read (1993), this test evaluates the vocabulary depth of L2 learners using word association tasks and semantic relationships between words. Qian (1999) obtained a reliability estimate of .91 for this test and Qian (2002) reported the split-half reliability for this measure to be .89.

The test has 40 recognition items which require the test takers to match four semantically-related words to target words using the relations of synonymy and collocations (see Figure 4). To decrease the possibility of blind guessing effect, the pattern of the responses has been designed in a way to follow three situations: "(a) the left and right boxes both contain two correct answers; (b) the left box contains one correct answer, and the right box contains three correct answers; and (c) the left box contains three correct answers, and the right box contains only one correct answer" (Qian, 2002, p. 524).

A   **correct**

easy free right true	answer dream enemy time

Figure 4. A WAT sample item.

Procedures

First, the OQPT was administered to test the language proficiency level of the learners and classify them into three groups of lower-intermediate (receiving 26-35 out of 60), upper-intermediate (receiving 36-45), and advanced (receiving 46-60). Second, the three tests of vocabulary size (i.e., the VST and its bilingual version as well as the VLT) were administered to the three groups of participants to find their vocabulary size. Finally, the participants took the WAT which measures the L2 students' vocabulary depth. The students took each test separately in different sessions, and the time allotted for each one was 30-45 minutes. Exact scoring method was used for all the five tests. Answer keys provided by the test developers were used for the

placement test as well as the three tests of vocabulary size. To score the test of vocabulary depth, the online scoring system at https://www.lextutor.ca/tests/associates/ was employed.

Results

Correlations between the L2 size and depth of vocabulary knowledge across the proficiency levels

The descriptive results that include the maximum possible score for each test, range, minimum, maximum, mean, and standard deviation are first presented for the three groups.

Table 1
Descriptive Statistics for All the Tests

Level	N	Test	MPS	Range	Min.	Max.	Mean	SD
Lower-intermediate	43	VLT	120	32.00	28.00	60.00	52.27	7.84
		VST Mono.	140	48.00	29.00	77.00	43.06	11.04
		VST Bi.	140	52.00	33.00	85.00	50.02	11.61
		WAT	100	45.00	29.00	74.00	50.69	10.66
Upper-intermediate	57	VLT	120	25.00	61.00	86.00	70.61	6.60
		VST Mono.	140	68.00	30.00	98.00	47.29	12.63
		VST Bi.	140	70.00	39.00	109.00	61.77	14.47
		WAT	100	33.00	52.00	85.00	67.78	8.87
Advanced	22	VLT	120	48.00	65.00	113.00	93.81	10.34
		VST Mono.	140	65.00	35.00	100.00	63.59	16.24
		VST Bi.	140	73.00	55.00	128.00	80.31	19.18
		WAT	100	42.00	46.00	88.00	73.36	10.68

Note: MPS = Maximum possible score; SD = Standard Deviation; Mono. = Monolingual; Bi. = Bilingual

Table 1 shows that the vocabulary size tests could differentiate between the three levels of language proficiency as more proficient students obtained higher scores. For instance, the lower-intermediate students' mean score on the VLT was 52.27, $(SD = 7.84)$, while the upper-intermediate learners' average score on this test was 70.61, $(SD = 6.60)$, and the advanced test takers received the highest score on the VLT $(M = 93.81, SD = 10.34)$. As shown in Table 1, the average score of the participants at the lower-intermediate level on the bilingual Persian VST $(M = 50.02, SD = 11.61)$ was higher than the English-only version $(M = 43.06, SD = 11.04)$. The average score of the upper-intermediate students on the two versions of the VST showed that the students performed better on the bilingual version $(M = 61.77 > 47.29)$. The mean score of the advanced-level students on the bilingual VST $(M = 80.31, SD = 19.18)$ was higher than the English-only version $(M = 63.59, SD = 16.24)$ as well.

The link between the size and depth dimensions of the L2 students' word knowledge with three language proficiency levels was addressed in the first research question. To answer this question, Pearson correlations were run to find the relationship between the scores of the test takers on the three tests of vocabulary size and the WAT.

Table 2
Pearson Correlations between the Three Tests of Vocabulary Size and Vocabulary Depth

Level	N	Test	WAT	*Sig.*	r^2
Lower-intermediate	43	VLT	.510**	.000	.260
		VST Monolingual	.412*	.006	.169
		VST Bilingual	.509**	.000	.259
Upper-intermediate	57	VLT	.399*	.002	.159
		VST Monolingual	.125	.353	.015
		VST Bilingual	.406*	.002	.164
Advanced	22	VLT	.345	.115	.119
		VST Monolingual	.156	.487	.024
		VST Bilingual	.314	.155	.098

Note: **Significant at $p = .001$; *Significant at $p = .01$

As shown in Table 2, the correlation coefficients between the lower-intermediate students' breadth and depth of lexical knowledge were statistically significant for the VLT ($r = .510$, $p < .001$), the VST ($r = .412$, $p < .01$) and its bilingual Persian version ($r = .509$, $p < .001$). The effect sizes of the correlations between the WAT and the VLT ($r^2 = .26$) and WAT and the bilingual VST ($r^2 = .25$) were large, while the effect size of the correlation between the WAT and the monolingual VST was medium ($r^2 = .16$) (Cohen, 1992). The overall results thus showed that the shared variance between the two aspects of size, as measured through the VLT and the bilingual Persian VST, and vocabulary depth was more significant than when the size was assessed by the monolingual VST.

As for the upper-intermediate students, the Pearson correlation between the breadth and depth of lexical knowledge was significant for the VLT ($r = .399$, $p < .01$) and the bilingual version of the VST ($r = .406$, $p < .01$). The relationship between the VST and scores on the WAT, however, was not significant ($r = .125$, $p > .05$). The effect size of the association between the VLT, bilingual Persian VST, and WAT was medium ($r^2 = .15$ and .16, respectively), while for the VST, the effect size was small ($r^2 = .01$) (Cohen, 1992).

For the advanced-level students, the association between the students' size and depth of lexical knowledge was not significant for any of the three tests. Pearson correlation between vocabulary depth and the VLT ($r = .345$, $p > .05$), the VST ($r = .156$, $p > .05$), and its bilingual Persian version ($r = .314$, $p > .05$) was not significant. The effect sizes of the corresponding correlations were medium to small as well ($r^2 = .11$, $r^2 = .02$, $r^2 = .09$). Overall, results indicated that students' performances on the three tests of vocabulary size and WAT, a word association test for depth of vocabulary, were not significantly related to each other for the advanced students.

Correlations between the higher-frequency vocabulary size and L2 vocabulary depth across the proficiency levels

The descriptive statistics for the second research question are displayed in Table 3. This includes the same information for the students' scores on the high-frequency-word levels of the VLT, the VST (from the 1,000- to the 7,000-word frequency level), and the bilingual VST (same as the monolingual VST). The results show that the test takers with higher language proficiency obtained better scores on the higher word-frequency sub-tests of the vocabulary size tests. Moreover, the students' performance on the bilingual VST was better than the corresponding scores on the monolingual VST.

Table 3
Descriptive Statistics for the Higher-Frequency Word Levels

Level	N	Sub-test	MPS	Range	Min.	Max.	Mean	SD
Lower-intermediate	43	High VLT	90	32.00	28.00	60.00	51.39	7.53
		High VST Mono.	70	35.00	25.00	60.00	39.27	9.07
		High VST Bi.	70	35.00	30.00	65.00	44.37	9.78
Upper-intermediate	57	High VLT	90	29.00	51.00	80.00	64.12	6.42
		High VST Mono.	70	35.00	30.00	65.00	44.43	9.32
		High VST Bi.	70	30.00	35.00	65.00	53.89	8.68
Advanced	22	High VLT	90	29.00	61.00	90.00	80.90	5.88
		High VST Mono.	70	39.00	31.00	70.00	55.50	11.55
		High VST Bi.	70	20.00	50.00	70.00	60.54	5.69

Note: MPS = Maximum possible score; SD = Standard Deviation; Mono. = Monolingual; Bi. = Bilingual

Pearson correlations were run to answer the second research question. Table 4 indicates that the link between the lower-intermediate test takers' higher-frequency vocabulary size and their vocabulary depth was statistically significant for the VLT ($r = .555$, $p < .001$), the VST ($r = .471$, $p < .01$), and the bilingual VST ($r = .481$, $p < .01$). The effect sizes of the shared variance were large for the VLT ($r^2 = .30$) and moderate for the English-only and bilingual version of the VST ($r^2 = .22$, $r^2 = .23$, respectively) (Cohen, 1992).

The relationship between the two aspects for the upper-intermediate students was moderate for all the tests of vocabulary size (VLT: $r = .392$, $p < .001$; VST: $r = .152$, $p > .05$; bilingual VST: $r = .383$ $p < .01$). The effect sizes of the shared variance were moderate as well. For the advanced participants, the correlations were not statistically significant ($p > .05$), and the effect sizes were small.

Table 4
Pearson Correlations between the Higher-Frequency Vocabulary Size and L2 Vocabulary Depth

Level	N	Test	WAT	*Sig.*	r^2
Lower-intermediate	43	High VLT	.555**	.000	.308
		High VST Mon.	.471*	.001	.222
		High VST Bi.	.481*	.001	.231
Upper-intermediate	57	High VLT	.392*	.003	.154
		High VST Mono.	.152	.260	.023
		High VST Bi.	.383*	.003	.147
Advanced	22	High VLT	.359	.101	.129
		High VST Mono.	.146	.517	.021
		High VST Bi.	.284	.200	.081

Note: **Significant at $p = .001$; *Significant at $p = .01$

Correlations between the lower-frequency vocabulary size and L2 vocabulary depth across the proficiency levels

The results for the last research question, as provided in Table 5, indicated that, like the students' scores on the overall as well as the higher word-frequency sub-tests of the VLT, the English-only and bilingual version of the VST, the lower word-frequency sub-tests differentiated between the three proficiency levels. In addition, the average score of the test takers on the bilingual Persian VST was more than the monolingual one.

Table 5
Descriptive Statistics for the Lower-Frequency Word Levels

Level	N	Sub-test	MPS	Range	Min.	Max.	Mean	SD
Lower-intermediate	43	Low VLT	30	4.00	0.00	4.00	.88	1.13
		Low VST Mono.	70	17.00	0.00	17.00	3.79	3.67
		High VST Bi.	70	20.00	0.00	20.00	5.65	3.86
Upper-intermediate	57	Low VLT	30	8.00	2.00	10.00	6.49	1.98
		Low VST Mono.	70	33.00	0.00	33.00	2.85	5.23
		Low VST Bi.	70	44.00	0.00	44.00	8.00	8.07
Advanced	22	Low VLT	30	24.00	4.00	28.00	12.90	6.60
		Low VST Mono.	70	27.00	3.00	30.00	8.22	6.57
		Low VST Bi.	70	58.00	5.00	63.00	19.77	15.15

Note: MPS = Maximum possible score; SD = Standard Deviation; Mono. = Monolingual; Bi. = Bilingual

Table 6 shows that the Pearson correlations between the lower word-frequency sub-tests of the vocabulary size tests and the WAT were not statistically significant for any of the three proficiency levels ($p > .05$). The only exception was the relationship between the two aspects, as measured through the bilingual Persian VST for the lower-intermediate participants ($r = .314$ $p < .05$) which had a moderate effect size ($r^2 = 09$).

Table 6
Pearson Correlations between the Lower-Frequency Vocabulary Size and L2 Vocabulary Depth

Level	N	Test	WAT	*Sig.*	r^2
Lower-intermediate	43	Low VLT	-.162	.300	.026
		Low VST Mon.	.074	.639	.005
		Low VST Bi.	.314*	.040	.098
Upper-intermediate	57	Low VLT	.057	.675	.003
		Low VST Mono.	.032	.811	.001
		Low VST Bi.	.260	.051	.067
Advanced	22	Low VLT	.220	.324	.049
		Low VST Mono.	.144	.523	.021
		Low VST Bi.	.291	.189	.084

Note: *Significant at $p = .05$

Discussion

This study investigated the correlations between the VLT, the VST, and its bilingual version as well as the L2 vocabulary depth for three language proficiency levels. Additionally, the relationship between the most and least frequent word lists of the three tests and the WAT were investigated. The results indicated that while the relationship between the two dimensions of word knowledge was significant for lower proficiency students, it was weak for higher proficiency levels. The findings showed that the relationship between the two variables became weaker as the proficiency level of the students increased. Moreover, the correlation analyses revealed that, while the relationship between the higher-frequency vocabulary size and L2 vocabulary depth was significant and large for the lower-intermediate students, it was moderate for the upper-intermediate and small for the advanced test takers. The corresponding relationship for the lower word-frequency bands was, however, insignificant and small for all the three proficiency levels.

The descriptive statistics provided evidence for the suitability of the vocabulary size tests for placement purposes since they could finely distinguish the low-proficient students from the high-proficient ones. This is in line with the arguments of earlier studies on vocabulary size that introduced the size tests as appropriate for placement purposes (e.g. Beglar, 2010; Beglar & Hunt, 1999; Janebi Enayat & Babaii, 2018; Laufer & Nation, 1999; Nation & Beglar, 2007).

The findings proved that the two aspects of vocabulary size and depth are interconnected. The results partially support previous studies which found higher correlations between vocabulary size and depth (e.g., Akbarian, 2010; Henriksen, 2008; Horiba, 2012). The significant relationships between the tests tapping into size of vocabulary and WAT further support the results of Zhang (2012) that the two dimensions are associated. The findings support Schmitt (2014, p. 941) who argued that "vocabulary size (as measured by the VLT) and lexical organization (as measured by the WAF) are strongly related to each other". The link between size and depth of word knowledge can be further explained using Meara and Wolter's (2004) argument that "[v]ocabulary size is not a feature of individual words: rather it is a characteristic of the test taker's entire vocabulary" (p. 87). They illustrate this by showing an interrelated network of words where adding new words (vocabulary size) leads to a better lexical organization (vocabulary depth).

The significant correlations, which had large to medium effect sizes, between the two dimensions of vocabulary, however, were true for the students at lower language proficiency levels and smaller vocabulary sizes, as this study found. This is in line with previous studies like Henriksen (2008) and Shimamoto (2000) which also reported that this interconnection varied across learners at different language proficiency and/or vocabulary sizes. These studies also showed that the scores on the tests of size and depth of word knowledge were more correlated for the participants who were less proficient. The lower association between the two dimensions of lexical knowledge for the students with a greater vocabulary size and higher language proficiency supports Schmitt (2014) that "for higher frequency words, and for learners with smaller vocabulary sizes, there is often little difference between size and a variety of depth measures" (p. 941).

The different correlations between the size and depth aspects of vocabulary across language proficiency levels, as this study found, are not consistent with some other studies. The results are not consistent with Nurweni and Read (1999) who found that the link between vocabulary size and lexical organization was closer for the learners with larger vocabulary sizes and higher language proficiency. Read (2004) reported the same opposite finding and Akbarian (2010) also showed that high-proficient ESP students' vocabulary size was more connected to their depth of vocabulary. The results, however, show that the association between vocabulary size and depth gets weaker as the students develop their language proficiency. This supports Akbarian (2010) that some other variables are involved in this relationship "so that we cannot take the size and depth of vocabulary knowledge as overlapping one another completely in the mental lexicon" (p. 400).

The weaker shared variance in the depth test, accounted for by the size tests for the advanced participants, further demonstrated that the depth test (i.e. the WAT) lags behind the size tests for higher proficiency levels. That is, while it can be used to test the students' depth of word knowledge at lower proficiency levels, it does not have the required level of difficulty to associate with the knowledge of lower frequency words, which is a common characteristic of more proficient students. Although this needs further investigation, it partially supports Milton (2009) that vocabulary depth tests like the WAT cannot measure the test takers' knowledge of semantic relationships among words.

The significant association between the higher-frequency vocabulary size and vocabulary depth for the low-proficient students on the one hand, and the lack of any correlation between the

lower-frequency vocabulary size and vocabulary depth for any of the proficiency levels on the other hand, were also indicative of the fact that vocabulary size and depth are correlated to the extent that the word-frequency levels are high and the proficiency levels of the students are low. This could thus further support Schmitt (2014, p. 941) that for lower-frequency words, "there is often a gap between size and depth, as depth measures lag behind the measures of size". In addition, this finding could be justified with reference to Greidanus and Nienhuis (2001) which examined the network knowledge of a group of EFL students and found that the test takers could identify more links among the higher-frequency words in comparison with the lower-frequency lexical items.

Conclusion

The results of the current study revealed that the three tests of vocabulary size had a significant relationship with vocabulary depth. However, the effect sizes of this relationship were large to moderate for the lower levels of language proficiency and small for the students with larger vocabulary sizes and higher language proficiency. The findings could clarify the argument raised by Schmitt (2014) that the strength of the correlation between size and depth of word knowledge across language proficiency is still inconclusive.

The findings of this study suggest some pedagogical implications. The findings indicate that the vocabulary size tests are appropriate means for placement purposes as they could differentiate between the lower- and higher-proficient students. These tests are easy to administer and score, and, taking into account the language proficiency level of the learners, the teachers and researchers could use different word-frequency bands.

The lack of a statistically significant association between the size and depth of word knowledge at higher language proficiency levels shows that teachers should pay more attention to the instruction of vocabulary depth for advanced students who have larger vocabulary size. In fact, this lack of correlation shows that, when students reach higher levels of language proficiency, their knowledge of vocabulary develops only in the size dimension and vocabulary depth becomes less significant. This provides an evidence that materials writers need to incorporate activities for the presentation, practice, and production of word associations as well (Brown, 2011). In fact, as Akbarian (2010) argues, vocabulary size and depth should not be taught separately, and they need to be taught in combination.

Similarly, test developers need to attend to the assessment of vocabulary depth, especially for higher proficient language learners. The inclusion of depth of word knowledge by test developers could, in turn, have a positive washback for the teaching and learning processes. Moreover, the results imply that the simultaneous assessment of vocabulary size and depth is suitable for lower proficiency levels but not the advanced students who have larger vocabulary size which, as this study found, is not associated with their depth of vocabulary knowledge.

Materials writers should also facilitate the instruction and assessment of vocabulary depth by including exercises that tap into the students' knowledge of word associations and provide them with tasks that require the use of "paradigmatic (superordinates, synonyms), syntagmatic (collocates), and analytic (words representing a key element of the meaning of the target word)" (Read, 2004, p. 221) relationships between words.

Notwithstanding the implications, this study had a number of limitations which should be addressed by future studies. First, this study only focused on the receptive measures of vocabulary size and depth. It is suggested that future studies investigate the same relationship for productive tests of vocabulary like the PVLT (Laufer & Nation, 1999) and Lex30 (Meara & Fitzpatrick, 2000). Moreover, the context of the study was limited to the university setting. Further studies could replicate this research with students from various contexts to increase the generalizability of the findings.

References

Akbarian, I. (2010).The relationship between vocabulary size and depth for ESP/EAP learners. *System, 38*(3), 391–401.

Alavi, S. M., & Akbarian, I. (2012). The role of vocabulary size in predicting performance on TOEFL reading item types. *System, 40*(3), 376–385.

Alharthi, T. (2020). Investigating the relationship between vocabulary knowledge and FL speaking performance. *International Journal of English Linguistics, 10*(1), 37-46.

Amirian, S.M.R., Mallahi, O., & Zaghi, D. (2015). The relationship between Iranian EFL learners' self-regulatory vocabulary strategy use and their vocabulary size. *Iranian Journal of Language Teaching Research, 3*(2), 29-46.

Avila Muñoz, A. (2017). The available lexicon: A tool for selecting appropriate vocabulary to teach a foreign language. *Iranian Journal of Language Teaching Research, 5*(1), 71-91.

Baba, K. (2009). Aspects of lexical proficiency in writing summaries in a foreign language. *Journal of Second Language Writing, 18*(3), 191–208.

Batty, O. (2007). Vocabulary depth in written and oral assessment. In K. Brandford-Watts (Ed.), *JALT 2006 conference proceedings* (pp.1100–1108). Tokyo: JALT.

Beglar, D. (2010). A Rasch-based validation of the Vocabulary Size Test. *Language Testing, 27*(1), 101–118.

Beglar, D., & Hunt, A. (1999). Revising and validating the 2000 word level and university word level vocabulary tests. *Language Testing, 16*(2), 131–162.

Brown, D. (2011). What aspects of vocabulary knowledge do textbooks give attention to? *Language Teaching Research*, *15*(1), 83–97.

Cheng, J., & Matthews, J. (2018). The relationship between three measures of L2 vocabulary knowledge and L2 listening and reading. *Language Testing, 35*(1), 3-25.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159.

Dabbagh, A. (2016). The predictive role of vocabulary knowledge in listening comprehension: Depth or breadth? *International Journal of English Language & Translation Studies, 4*(3), 1-13.

Dabbagh, A., & Janebi Enayat, M. (2019). The role of vocabulary breadth and depth in predicting second language descriptive writing performance. *The Language Learning Journal, 47*(5), 575-590.

Geranpayeh, A. (2003). A quick review of the English Quick Placement Test. *Research Notes, 12*, 8–10.

Greidanus, T., & Nienhuis, L. (2001). Testing the quality of word knowledge in L2 by means of word associations: Types of distractors and types of associations. *Modern Language Journal, 85*(4), 567–77.

Gyllstad, H. (2007). *Testing English collocations*. Lund: Lund University, MediaTryck.

Haastrup, K., & Henriksen, B. (2000). Vocabulary acquisition: Acquiring depth of knowledge through network building. *International Journal of Applied Linguistics, 10*(2), 221–240.

Henriksen, B. (1999).Three dimensions of vocabulary development. *Studies in Second Language Acquisition, 21*(2), 303–317.

Henriksen, B. (2008). Declarative lexical knowledge. In D. Albrechtsen, K. Haastrup, & B. Henriksen (Eds.), *Vocabulary and writing in a first and second language* (pp. 22–66). Basingstoke, UK: Palgrave Macmillan.

Horiba, Y. (2012). Word knowledge and its relation to text comprehension: A comparative study of Chinese- and Korean-speaking L2 learners and L1 speakers of Japanese. *Modern Language Journal, 96*(1), 108–121.

Ishii, T., & Schmitt, N. (2009). Developing an integrated diagnostic test of vocabulary size and depth. *RELC Journal, 40*(1), 5-22.

Janebi Enayat, M., & Babaii, E. (2018). Reliable predictors of reduced redundancy test performance: the interaction between lexical bonds and test takers' depth and breadth of vocabulary knowledge. *Language Testing, 35*(1), 121-144.

Kaivanpanah, S., & Zandi, H. (2009).The role of depth of vocabulary knowledge in reading comprehension in EFL context. *Journal of Applied Sciences, 9*(4), 698–706.

Karami, H. (2012). The development and validation of a bilingual version of the Vocabulary Size Test. *RELC Journal, 43*(1), 153–167.

Laufer, B., & Nation, P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing, 16*(1), 33–51.

Matthews, J. (2018). Vocabulary for listening: Emerging evidence for high and mid-frequency vocabulary knowledge. *System*, *72*, 23-36.

Matthews, J., & Cheng, J. (2015). Recognition of high frequency words from speech as a predictor of L2 listening comprehension. *System, 52*(1), 1–13.

Meara, P., & Fitzpatrick, T. (2000). Lex30: an improved method of assessing productive vocabulary in an L2. *System 28*(1), 19-30.

Meara, P., & Wolter, B. (2004). V-Links: Beyond vocabulary depth. In Albrechtsen, D., Haastrup, K., & Henriksen, B. (Eds.), *Writing and vocabulary in foreign language acquisition* (pp. 85-96). Denmark: Museum Tusculanum Press.

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

Nassaji, H. (2006). The relationship between depth of vocabulary knowledge and L2 learners' lexical inferencing strategy use and success. *Modern Language Journal, 90*(3), 387–40.

Nation, I. S. P. (1983). Testing and teaching vocabulary. *Guidelines, 5*(1), 12-25.

Nation, I. S. P. (2001). *Learning vocabulary in another language.* Cambridge: Cambridge University Press.

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

Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Japan Association for Language Teaching, 31*(7), 9-12.

Noro, T. (2002). The roles of depth and breadth of vocabulary knowledge in reading comprehension in EFL. *ARELE*, *13*, 71–80.

Nurweni, A., & Read, J. (1999). The English vocabulary knowledge of Indonesian university students. *English for Specific Purposes*, *18*(2), 161–175.

Qian, D. (1999). Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension. *Canadian Modern Language Review, 56*(2), 283-307.

Qian, D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: an assessment perspective. *Language Learning, 52*(3), 513–536.

Qian, D., & Schedl, M. (2004). Evaluation of an in depth vocabulary knowledge measure for assessing reading performance. *Language Testing, 21*(1), 28-52.

Read, J. (1993).The development of a new measure of L2 vocabulary knowledge. *Language Testing, 10*(3), 355–371.

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

Read, J. (2004). Plumbing the depth: How should the construct of vocabulary knowledge be defined? In P. Bogaards, & B. Laufer (Eds.) *Vocabulary in a second language* (pp. 209–227). Amsterdam: Benjamins.

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual.* Houndmills, Basingstoke, Hampshire; New York: Palgrave Macmillan.

Schmitt, N. (2014). Size and depth of vocabulary knowledge: What the research shows. *Language Learning, 64*(4), 913–951.

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing, 18*(1), 55-88.

Shimamoto, T. (2000). An analysis of receptive vocabulary knowledge: Depth versus breadth. *JABAET, 4*, 69–80.

Shiotsu, T., & Weir, C. (2007). The relative significance of syntactic knowledge and vocabulary breadth in the prediction of reading comprehension test performance. *Language Testing, 24*(1), 99–128.

Stæhr, L. S. (2008) Vocabulary size and the skills of listening, reading and writing. *Language Learning Journal, 36*(2), 139–152.

Stæhr, L. S. (2009). Vocabulary knowledge and advanced listening comprehension in English as a foreign language. *Studies in Second Language Acquisition, 31*(4), 577–607.

University of Cambridge: *ESOL Examinations (2004). Quick Placement Test* (Version 1). Oxford University Press.

Zhang, B., & Anual, S. B. (2008). The role of vocabulary in reading comprehension: The case of secondary school students learning English in Singapore. *RELC Journal 39*(1), 51–76.

Zhang, B., & Li, C. (2011). Classification of L2 vocabulary learning strategies: Evidence from exploratory and confirmatory factor analysis. *RELC Journal, 42*(2), 141–154.

Zhang, D. (2012). Vocabulary and grammatical knowledge in L2 reading comprehension: A structural equation modeling study. *Modern Language Journal, 96*(4), 558-575.

Mostafa Janebi Enayat is an Assistant Professor of Applied Linguistics at University of Maragheh, Maragheh, Iran. His publications have appeared in accredited international journals, such as *Language Testing*, *The Language Learning Journal*, and *Innovation in Language Learning and Teaching*. His main research interests are L2 Vocabulary Assessment and Critical Discourse Analysis.

Seyed Mohammad Reza Amirian is an Associate Professor of Applied Linguistics at Hakim Sabzevari University where he teaches undergraduate and post-graduate courses. His research interests are language testing and assessment, assessing language skills, test fairness and DIF, vocabulary instruction, corrective feedback, and teacher education.