

Research on the correlation between vocabulary depth knowledge and reading proficiency of English majors

Ke Du

School of Foreign Studies, Nanjing Forestry University, Nanjing, Jiangsu, 210037, China

Keywords: Vocabulary depth; reading comprehension; reading ability

Abstract: Since vocabulary knowledge has been regarded as a key predictor of second language learners' reading comprehension, vocabulary acquisition and related works of lexical pedagogy have always been heatedly discussed within the fields of applied linguistics and second language learning and teaching. This study seeks to show build a better understanding of the relationship between second language learners' vocabulary depth and their reading proficiency. Through the descriptive and correlation analysis of experimental tests, the study demonstrated that vocabulary depth influences learners' reading comprehension and linguistic ability, which highlights the significance of lexical pedagogy for enhancing learners' reading proficiency.

1. Introduction

Vocabulary acquisition is central to any language acquisition process, native or non-native. Wilkins once pointed out that "Without grammar, little can be conveyed; without vocabulary, nothing can be conveyed." A number of related research has indicated that vocabulary is the basis of learning a language and also offers a great challenge to language learners, especially in enhancing learners' second language reading ability [1-4] If the acquisition of language is compared to a building, vocabulary is always the cornerstone of this building. Schmitt & Clapham also proposed that vocabulary is an important "building material" of language, which highlights the lexical nature of language learning. It can be seen that the study and mastery of vocabulary are indispensable and extremely important to language acquisition. English teaching has received great attention in China's current education, and English ability is mainly tested by examination, which is divided mainly into six interlinking aspects: speaking, listening, grammar, gesture-type, reading, and writing. Taking English learning as an example, for most language learners, the objection to building stronger vocabulary ability is to get better results in English tests, which usually presents a high demand on test-takers' lexical variation. However, it actually emphasizes the breadth of vocabulary knowledge that people are familiar with, while the depth and sophistication of vocabulary knowledge are often overlooked. In fact, the breadth and depth of vocabulary knowledge are equally important to vocabulary learning and language pedagogy [2]. Therefore, the investigation of the internal relationship and mutual influence between learners' vocabulary depth and the six aspects of English proficiency tests may play an important role in the improvement of learners' English proficiency [5].

Within the field of second language acquisition, the relationship between learners' lexical

knowledge, which is a specific form of knowledge, and language ability has been widely discussed [6]. The ability of English reading serves as an important index to examine the quantity and quality of second language input of English learners. As a result, scores of reading comprehension in various English tests often take a large proportion. Taking CET-6 as an example, the total score of the reading part is 249, accounting for 35 % of the total score of the exam. The central concept, depth of vocabulary knowledge refers to the second language learners' understanding, familiarity, and mastery of all the features of one word, rather than the facial process of knowing its "symbol-meaning". Recent years have witnessed a great number of foreign studies focusing on the relationship between vocabulary depth knowledge and reading and writing. Studies concerning the breadth of second language learners' vocabulary knowledge are abundant [7]. For example, previous studies have shown that the breadth of second language learners' vocabulary knowledge contributes to the development of learners' comprehensive language proficiency, among which only a limited number of studies touch on the benefits of both the breadth and depth of learners' lexical competence[8,9]. And the studies that give special attention to the depth of learners' vocabulary knowledge are relatively limited. In terms of related research in China, few domestic scholars have explored the relationship between vocabulary size and reading skills despite some studies discussing the relationship between language comprehensive ability and vocabulary depth knowledge.

The present study attempts to explore the influence of the depth of vocabulary knowledge on the reading proficiency of Chinese English learners in the environment of English as a foreign language (EFL), taking English majors as the experimental participants. More particularly, it aims to answer the following two questions: (1) What is the relationship between second language learners' depth of vocabulary knowledge and their reading proficiency? How do they correlate with each other? (2) To what extent can reading vocabulary size explain the difference in reading comprehension scores?

2. Methodology

2.1 Research Participants

The participants of this study are 38 junior English majors at a university in Nanjing, including 31 female students and 7 male students. Most of them are from Jiangsu Province and have adequate English proficiency. Thirty-five valid data were obtained after the invalid test volume was removed.

2.2 Research Design

The depth of vocabulary knowledge in this study was measured by the Word Associate Test (with a reliability of 0.93) designed by Read. 40 target words are designed for the experimental test, and each question contains two boxes both on the left and right, which test respectively the vertical aggregation relationship of the target words (near-synonyms and polysemous items of the target words), and the horizontal combination relationship (collocation, with a total of four correct choices). The table 1 shows an example of the test word "sound". Words in the left-hand box that have a similar meaning to "sound" are "logical", "healthy" and "solid" respectively, and the wanted word in the right-hand box that can be used with sound is "sleep". Usually, the distribution of correct answers may also be just one on the left, three on the right, or both two on each side. The irregular distribution of the right answer can effectively reduce participants' guessing and invalid data. Each question of this test is worth 2.5 marks, with no deductions for incorrect answers. This test is scored out of 100 points.

The reading level test in this study was selected from reading comprehension in TEM 4 (Test for English majors Grade 4), which is a national English professional test providing a relatively

objective and accurate measure of the actual English proficiency of Chinese English majors. This study excerpts the multiple-choice questions in the 2019 and 2020 TEM 4 reading comprehension part as the test instrument to test participants' reading comprehension ability. This test consists of six articles and 20 multiple-choice questions, each worth 2.5 points, for a total of 50 points.

Table 1: The example of the test word “sound”

left	right
Logical; healthy; bold; solid	Snow; temperature; sleep; dance

This study used qualitative research to explore the nature of the depth of participants' vocabulary knowledge and the relationships between it and the development of participants' reading comprehension ability. To be more specific, this study, applying the statistical method, centers on numerical scores with the help of a more rigorous design and standardized procedures to investigate the variation of the variables and their relationships [10].

2.3 Data Collection

The data was collected in two separate sessions. Before the test, participants were explicitly informed that all data were used for study purposes and that the tests were conducted anonymously. Data collection is carried out in two stages:

a. Vocabulary depth knowledge test: The participants were asked to finish a vocabulary depth test within a limited time on Wenjuanxing (a platform providing functions equivalent to Amazon Mechanical Turk). After the test, the researchers checked the participants' final scores and recorded the data.

b. Reading level test: Participants were given a reading comprehension test. They were asked to finish the test within the time limit on Wenjuanxing. According to the scores, the researchers graded the papers submitted by the participants, obtained the final scores, and recorded them carefully [10].

The raw data are show as Table 2.

Table 2: The example of the test word “sound”

Depth of vocabulary knowledge	78, 84, 70, 70, 82, 80, 86, 82, 72, 80, 82, 70, 66, 80, 88, 82, 74, 74, 80, 70, 66, 68, 90, 76, 80, 78, 72, 82, 74, 88, 86, 70, 76, 76, 78
Reading proficiency	40, 42.5, 27.5, 30, 35, 35, 37.5, 35, 30, 37.5, 40, 27.5, 25, 30, 40, 37.5, 30, 30, 32.5, 30, 27.5, 27.5, 45, 40, 35, 37.5, 32.5, 37.5, 35, 35, 37.5, 27.5, 30, 35, 37.5

2.4 Data processing

The two variables in this study are continuous variables, and their data values are independent. The data was processed by Pearson correlation analysis. SPSS 25.0 was used to draw scatterplots and to conduct Shapiro-Wilk normal distribution test and descriptive statistical analysis. Later, the depth of vocabulary knowledge of the participants was described, followed by the Pearson correlation analysis

Pearson correlation analysis, designed by Karl Pearson, is mainly a parametric test. Pearson correlation analysis requires the following five statistical assumptions: (1) Observations must be distance or ratio data, that is, the relevant variable is a continuous variable; (2) The observations are independent, and there are no outliers; (3) The data of the two variables obey (or approximate obey) normal distribution; (4) The homoscedasticity is concerned; (5) Linearity is concerned. Besides, the independence of observations means that there is no correlation between observations for different subjects.

3. Discussion

3.1 Descriptive Statistical Analysis

The data of the two variables in this study are both continuous data, so Pearson correlation analysis is primarily considered. Before conducting it, it is necessary to first test the statistical hypothesis of the analysis and conduct a descriptive statistical analysis. The Shapiro-Wilk normal distribution test and the box plot method were used to check whether the two variable data obeyed or approximately obeyed the normal distribution, the characteristics of the data distribution and whether there were outliers among the data.

The operation steps of SPSS 25.0 to describe the variation in the deep vocabulary knowledge and reading proficiency are as follows.

(1) *To describe the variation in the deep vocabulary knowledge:*

First, SPSS 25.0 was opened and the two sets of data were input in sequence and the variables were defined accordingly. Then chose “scale” in the “change” column and enter the main dialog box in the sequence of “Analyze->Descriptive Statistics->Explore...”

Second, in the main dialog box, the left blanket variable “vocabulary depth knowledge” was typed into the Dependent List field. All options in the window adopted the default mode of SPSS.

Third, clicked on “Statistics...”, and “Descriptives” was selected.

Fourth, clicked on “Plots...” and select “Normality plots with tests”. The remains were carried the default SPSS method. Then “Continue” was clicked to return to the main exploratory analysis dialog.

Fifth, “OK” was clicked to complete the operation.

(2) *To describe the variation in reading proficiency:*

First, entered the main dialog box in the sequence of “Analyze->Descriptive Statistics->Explore...” as the above steps.

Second, in the main dialog box, the left blanket variable “reading proficiency” was typed into the Dependent List field. All options in the window adopted the default mode of SPSS.

Third, clicked on “Statistics...”, and “Descriptives” was selected.

Fourth, clicked on “Plots...” and select “Normality plots with tests”. The remains were carried the default SPSS method. Then “Continue” was clicked to return to the main exploratory analysis dialog.

Fifth, “OK” was clicked to complete the operation.

The results of each procedure are show as Tables 3 and 4, respectively.

Table 3: Description of the deep vocabulary knowledge

	Descriptives	Statistic	Std. Error	
Depth of vocabulary knowledge	Mean	77.43	1.099	
	95% Confidence Interval for mean	Lower Bound	75.20	
		Upper Bound	79.66	
	5% Trimmed Mean	77.41		
	Median	78.00		
	Variance	42.252		
	Std. Deviation	6.500		
	Minimum	66		
	Maximum	90		
	Range	24		
	Interquartile Range	10		
	Skewness	0.015	0.398	
	Kurtosis	-0.838	0.778	

Table 4: Description of reading proficiency

Descriptives		Statistic	Std. Error	
Reading Proficiency	Mean	34.07	0.839	
	95% Confidence Interval for mean	Lower Bound	32.37	
		Upper Bound	35.78	
	5% Trimmed Mean	35.97		
	Median	35.00		
	Variance	24.664		
	Std. Deviation	4.966		
	Minimum	25		
	Maximum	45		
	Range	20		
	Interquartile Range	8		
	Skewness	0.097	0.398	
	Kurtosis	-0.789	0.778	

As shown in the description tables, the average scores of the vocabulary depth knowledge test and reading proficiency test are 77.43 and 34.07 respectively, and the skewness value and kurtosis value are close to 0, indicating that the data of the two variables may be approximately subject to normal distribution.

As shown in the box-plots, the median of the data of the vocabulary depth knowledge test and reading proficiency test is slightly away from the center of the box, with relatively average distributions. And no obvious outliers are shown in the Figures 1 and 2.

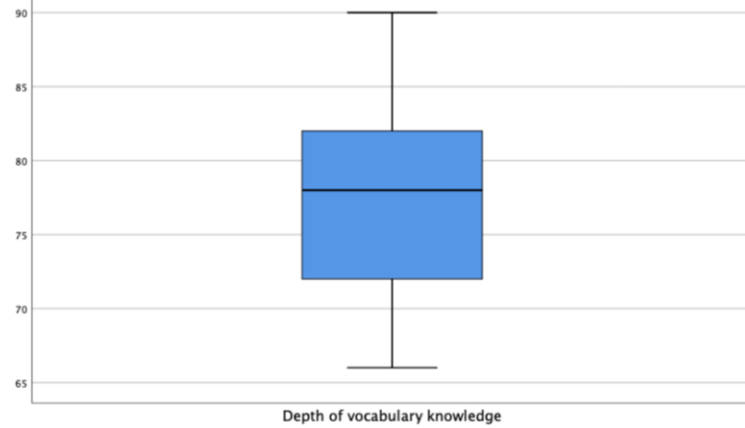


Figure 1: Box-plot of the deep vocabulary knowledge

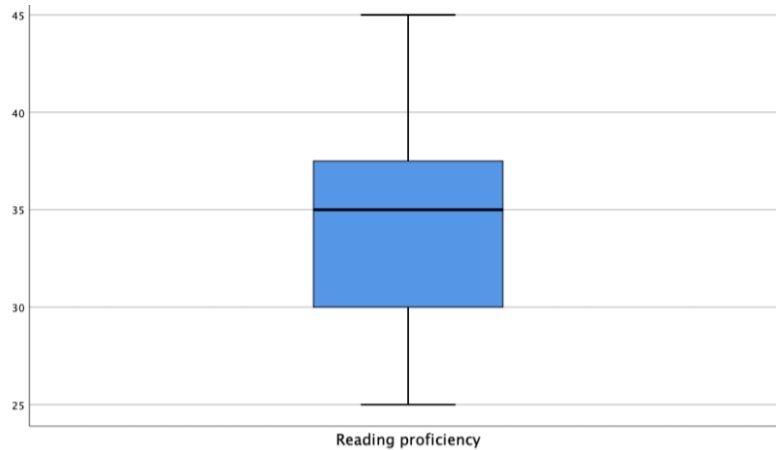


Figure 2: Box-plot of reading proficiency

As shown in the normality test chart (figure 3), among the data of “deep vocabulary knowledge”, $W=0.97$ ($p=0.37>0.05$), which accepted the null hypothesis, and the W value was close to 1, approximately following the normal distribution. And the data of the “reading proficiency” indicated $W=0.95$ ($p=0.12>0.05$), which accepted the null hypothesis, and the W value was close to 1, which also approximately followed the normal distribution.

Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Reading proficiency	.165	35	.017	.951	35	.122
Depth of vocabulary knowledge	.111	35	.200*	.967	35	.367

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Figure 3: Normality test

3.2 Pearson Correlation Analysis

Before Pearson correlation analysis, the scatterplot was used to explore whether the scores of the depth of vocabulary knowledge test and reading proficiency test were linearly related, whether there were outliers, and whether the variance was homogeneous, etc.

The operation steps of SPSS 25.0 to describe the correlation between the deep vocabulary knowledge and reading proficiency are as follows.

First, went to the graph dialog in the order of “Graphs>Chart Builder”. Then dragged a Simple Scatterplot with the fit line into the “Chart preview” area by selecting “Simple Scatter With Fit Line” from the “Choose From” list.

Second, the “deep vocabulary knowledge” was dragged from the “Variables” column into “X-Axis?”. Next, in the specified rectangular box, “reading level” was dragged into the “Y-Axis?” of the “Chart preview” area.

Third, in the Element Properties window, “Title 1” in the “Edit Properties of” panel and “None” were selected in order.

Fourth, clicked “Chart Appearance”, tick “Use custom color”, “border”, and the “grid lines settings” in the “Editor Appearance” column. Then “Inner” in “Frame” column was selected.

Fifth, clicked “OK” and finished the proceeding operation.

Sixth, double-clicked the graph generated by SPSS 25.0 to enter the “Chart Editor” window. Then the “Fit Lines at Subgroups” was selected from the drop-down menu in the “Element”. The SPSS default was accepted. And next removed the default “Attach label to line” option of the program. After that, the “Apply” and “Close” buttons were clicked.

Seventh, any value on the vertical axis is double-clicked and the “Labels&Ticks” in the activated properties window was selected. Checked the option “Displayticks” in the “Major Ticks” and “Minor Ticks” panels separately with SPSS default scale Style: Outside. Next, clicked “Apply” and then “Close”. Again, any value on the horizontal axis was double-clicked, and “Labels&Ticks” in the activated properties window was selected. Checked the option “Displayticks” in the “Major Ticks” and “Minor Ticks” panels with SPSS default scale Style: Outside, and then clicked “Apply” and “Close”.

Finally, the “Chart Editor” was closed and the data processing was finished. The results are shown in the Figure 4.

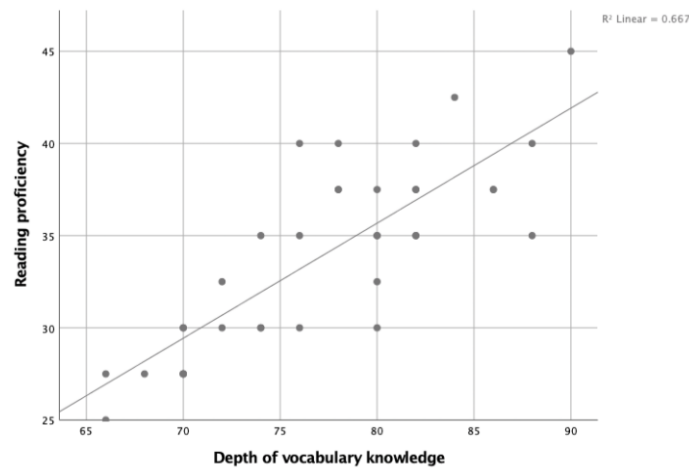


Figure 4: Scatterplot of deep vocabulary knowledge and reading proficiency

The predictive power $R^2=0.515$ in the scatterplot indicates that learners' deep vocabulary knowledge can explain or predict 66.7% of the variation in their reading proficiency. Besides, it appears that with the increase of the depth of their vocabulary knowledge, reading proficiency is also on the rise accordingly. In this study, the data of the two variables were linearly related, no obvious outliers were shown, and the variance was uneven.

To sum up, the data in this study basically satisfy the hypothesis of Pearson correlation analysis mentioned earlier, which is suitable for Pearson correlation analysis. The detailed SPSS operation steps are as follows:

First, the "Bivariate Correlation" dialog box was opened through the "Analyze->Correlate->Bivariate..." procedure. Then the two variables "deep vocabulary knowledge" and "reading proficiency" were typed in the left column into the variables column on the right area. The other options default to Pearson correlation coefficient and two-tail test.

Second, clicked the button "Option" and entered the sub-dialog box of "Options", and checked the "Means and standard deviations" in the statistics panel. The other options were set as default.

Third, clicked "Continue" to return to the relevant analysis column main dialog box, and clicked "OK" to end the whole operation process.

To give a better overview of the data, Table 5 was provided.

Table 5: Descriptive statistical analysis and Pearson Correlation Analysis

Descriptive Statistics			
\	Mean	Std. Deviation	N
Reading proficiency	34.07	4.966	35
Depth of vocabulary knowledge	77.43	6.500	35
Correlations			
\		Reading Proficiency	Depth of vocabulary knowledge
Reading Proficiency	Pearson Correlation	1	0.817**
	Sig. (2-tailed)		0.000
	N	35	35
Depth of vocabulary knowledge	Pearson Correlation	0.817**	1
	Sig.(2-tailed)	0.000	
	N	35	35

** . Correlation is significant at the 0.01 level (2-tailed).

As shown in the table, each variable has a Pearson correlation coefficient of 1 (perfect correlation) with itself without the necessity of reporting the significance test results of the two-tail

test. Therefore, it can be concluded that the depth of participants' vocabulary knowledge was strongly correlated with their reading proficiency ($r=0.81$, $p=0.000<0.01$), i.e., deep vocabulary knowledge does have a positive effect on the participants' reading comprehension.

4. Discussion

Above all, the depth of vocabulary knowledge and reading comprehension are important indicators to differentiate between English majors and non-English majors. Therefore, the study yields insights into the enhancement of lexical and reading proficiency of Chinese English majors. First, as higher-level language learners, English majors should not only focus on the expansion of their vocabulary bank in daily learning but also pay attention to the improvement of their depth of vocabulary knowledge as well as their overall language ability and comprehensive ability through reading exercises. Second, English majors are supposed to realize that the knowledge they learn in the classroom is usually the tip of the iceberg, which aims to elicit students' interest in learning English. English majors should take the initiative to strengthen deep learning after class rather than be satisfied with the current language level. In addition, the conclusion of this study has certain implications for English reading and vocabulary teaching. When teachers know that the vocabulary of English majors is basically up to the standard, they are supposed to apply multiple activities to improve students' lexical proficiency to a higher level through teaching synonyms, antonyms, and collocations, as well as guiding the students to grasp the deep meaning of words in classroom context.

In general, vocabulary knowledge can be enlarged and enriched through a wide range of reading comprehension. Reading comprehension can be facilitated by a profound and solid foundation of vocabulary knowledge. On the basis of their general relationship, the present study attempts to explore in detail the effects of vocabulary knowledge on reading comprehension. In this study, if not a causal connection, at least some strong correlative pieces of evidence are expected to be found. If a substantial psychometric relationship is found, then we can propose that an important way to improve second language learners' reading comprehension is to increase their vocabulary. If so, it will have a great implication for our English reading teaching that we should put more emphasis on basic language knowledge teaching, particularly vocabulary teaching to improve students' reading performance.

5. Conclusion

The study applied Pearson correlation analysis and answered the research questions. It is found that there was a statistically significant correlation between the depth of vocabulary knowledge and reading proficiency of English majors by showing a high positive correlation between the two variables ($r=0.81>0.05$). Given that the study has demonstrated that the higher the level of second language learners' depth of vocabulary knowledge, the higher their reading proficiency and comprehension ability. Moreover, the research data can support the fact that lexical depth knowledge can explain or predict 66.7% of the variation in second language learners' reading ability. Based on the results, the study, therefore, suggests that lexical pedagogy focusing on learners' vocabulary depth may enhance their second language reading comprehension.

References

- [1] Brooks G., Clenton J., & Fraser S. *Exploring the Importance of vocabulary for English as an additional language learners' reading comprehension. Studies in Second Language Learning and Teaching*. 2021, 11(3), 351-376.
- [2] Jun Zhang L., & Bin Anual S. *The role of vocabulary in reading comprehension: The case of secondary school students learning English in Singapore. RELC Journal*, 2008, 39(1), 51-76.

- [3] Lestari M., & Wahyudin A. Y. *Language learning strategies of undergraduate EFL students. Journal of English Language Teaching and Learning*, 2020, 1(1), 25-30.
- [4] Wilkins E. J. *Foreign Language Instruction and Career Preparation: A Selected, Annotated Bibliography (No. 51). ERIC Clearinghouse on Languages and Linguistics, Center for Applied Linguistics*, 1997.
- [5] Nurweni A. & Read J. *The English vocabulary knowledge of Indonesian university students. English for Specific Purposes*, 1999, 18(2), 15-27.
- [6] Afshari S., & Tavakoli M. *The relationship between depth and breadth of vocabulary knowledge and Iranian EFL learners' listening comprehension. International Journal of Research Studies in Language Learning*. 2017, 6(3), 13-24.
- [7] Mehrpour S, Razmjoo S A, Kian P. *The Relationship between Depth and Breadth of Vocabulary Knowledge and Reading Comprehension among Iranian EFL Learners By [J]. Two Quarterly Journal of English Language Teaching and Learning University of Tabriz*, 2011, 2(222): 97-127.
- [8] David A. *Vocabulary breadth in French L2 learners. Language Learning Journal*, 2008, 36(2), 167-180.
- [9] Leontjev D., Huhta A., & Tolvanen A. *L2 English vocabulary breadth and knowledge of derivational morphology: One or two constructs? Language Testing*, 2023, 40(2), 300-324.
- [10] Zhang D. *Vocabulary and grammar knowledge in second language reading comprehension: A structural equation modeling study. The Modern Language Journal*, 2012, 96(4), 558-575.