

Investigating the Correlation Between Vocabulary Knowledge and Reading Comprehension in Expository Text of Senior High School Students

Aisyah Fairuz Zahra* & Fauris Zuhri

English Education Department, Faculty of Languages and Arts, State University of Surabaya, Jl. Lidah Wetan, Lidah Wetan, Kec. Lakarsantri, kota Surabaya, Jawa Timur 60213. Indonesia

*Corresponding Author: fairuzahraa@gmail.com

Article History

Received : August 16th, 2025

Revised : September 17th, 2025

Accepted : October 08th, 2025

Abstract: Reading is one of the essential skills to develop their language learning. For senior high students, they need to comprehend the text, particularly expository text. However, many students struggling with reading comprehension due to their limited vocabulary knowledge. Therefore, this research aimed to investigate the senior high school students' vocabulary knowledge in relation with reading comprehension, specifically in the context of expository text. A quantitative correlational research design was employed in this study, involving 65 students in the tenth grade from a state school in Surabaya, selected through convenience sampling. The researcher administered two tests, such as The Updated Vocabulary Level Test (TUVLT) by Paul Nation and reading comprehension test focusing on expository text to the participants involved in this research. The results revealed that the Pearson Product-Moment correlation value was 0.792 (r value). It showed that students' vocabulary knowledge and their expository text reading comprehension were positive and strongly correlated. Drawing on these findings, it can be concluded that students with higher vocabulary knowledge are more likely to perform greater in reading comprehension tasks.

Keywords: Expository Text, Reading Comprehension, Vocabulary Knowledge

INTRODUCTION

Reading is recognized as an essential English skill that second language (L2) learners need to develop for academic success. It is strongly connected to the reader's understanding and effective engagement with a text. A key aspect of reading is the capacity to decode words efficiently and precisely, which contributes to reading fluency. This skill enables individuals to expand their knowledge, broaden their perspectives, and stimulate their thinking. Since the majority of information, academic publications, and learning materials are conveyed in written form (Baker, 2018), reading skills are regarded as a priority in the learning process. Thus, reading remains crucial for students across all educational stages, beginning at the onset of schooling and continuing throughout their lives (Castillo et al., 2024). Nevertheless, as reading has a direct impact on both academic achievement and future employment prospects, it becomes one of the abilities that senior high school students should learn. On the other hand, the students will face more complex texts that

require critical thinking, which includes analysis, interpretation, and evaluation (Perfetti & Stafura, 2014). Students with strong reading skills are tend to obtain better results in reading comprehension and deal with the academic demands faced in higher education successfully. According to research by Dakhi (2019) the more frequently students engaged in reading, the better they grasped the meaning of individual English words. This was demonstrated by the results of several students who comprehended the material they were reading.

To cultivate a deeper comprehension of reading, students need to meet certain requirements. One of the key elements that affect students' reading ability is vocabulary, which provides the essential basis for comprehending written texts. Research has shown that the reading comprehension of senior high school students is greatly impacted by their vocabulary knowledge, as well as their capacity to infer meaning from context and engage in critical thinking (Alshehri, 2024). This statement further validated by (Bombo et al., 2022) who assert that

vocabulary knowledge is one of the primary factors affecting reading comprehension.

For the purpose of improving both learning outcomes and reading abilities, students need to be better understanding the vocabulary. Students who master the vocabulary are believed to be able to maintain their reading process, which enabling them to interpret texts effectively by relying on vocabulary as a key component of reading comprehension (Gallagher et al., 2019). The correlation between these variables highlights the importance of teachers emphasizing vocabulary development in their instructional methods, as it significantly affects students' reading comprehension. In the context of English learning, readers are unable to completely grasp a passage if they do not understand the meanings of the words it contains. The lack of recognizing the word's meaning can hinder their ability to process key information, which is essential for understanding the passage as a whole.

Numerous research had been carried out and provide valuable insights about the correlation between vocabulary knowledge and reading comprehension. For example, Manihuruk (2020) studied the relationship between EFL students' reading comprehension and vocabulary knowledge. According to the study, senior high school students' reading comprehension skills and vocabulary knowledge have a moderately good relationship. Moreover, this is support with research by Masrai (2019) which emphasize the need for educational strategies to incorporate vocabulary development in teaching and learning, as it plays a crucial role in supporting L2 reading comprehension within English language contexts.

According to the prior research, there are several numbers of limitations. Since these studies explored the connection between vocabulary knowledge and reading comprehension but did not clarify which text types were applied in the research. Therefore, it is necessary to investigate the particular texts students engage with, as this factor could affect the relationship. In particular, expository texts are considered appropriate for studies with senior high school students because they mirror the academic challenges these learners typically face. In addition, previous research did not account for the level of text comprehension. Consequently, the current study employed a reading comprehension passage aligned with CEFR level B2, which corresponds to an intermediate

English proficiency level and appropriately represents senior high students' reading abilities. Moreover, previous research often employed a broad range of vocabulary levels, making it challenging to pinpoint the specific impact of mid-frequency vocabulary on comprehension. Therefore, this research addressed this gap by concentrating on the 2000- and 3000-word levels, as these are regarded as more appropriate for helping senior high school students comprehend expository texts more effectively.

Based on the background described above, this study addresses the problem of senior high school students who appear to have difficulty understanding vocabulary, particularly in academic contexts, which affects their academic performance, especially their reading skills. Hence, the researcher developed assessments to explore the relationship between students' vocabulary mastery and their reading comprehension, with the purpose of gathering pertinent data and insights. Accordingly, the aim of this study is to determine whether senior high students at a Surabaya school have a positive relationship between their vocabulary knowledge and their ability to comprehend expository texts.

METHODS

Based on the purpose of this study, the researcher measured and analyzed the data by using a quantitative method. As stated by Creswell (2012), quantitative research is research that works with numbers (scores and grades, rank, or frequency). Specifically, correlational research design was applied to find out the correlation between variables involved. The researcher conducted the study to the students in the 10th grade at a Surabaya state high school, during the second semester of the 2024/2025 academic year, specifically in March 2025. The participant of this study is determined by using convenience sampling, which known as a technique used to select population members who are readily available and reachable during a specific period (Etikan, 2016). Since the population of this study consisted of all tenth-grade students at a state senior high school in Surabaya. A total sample of 65 students were selected as participants in this research that had already passed an expository text material in their English class by using convenience sampling as mentioned before. The two selected classes with 65 students in total were used to represent the

larger population, as they have similar academic backgrounds and learning experiences to other tenth graders.

To investigate and explore students' mastery of vocabulary within their expository text reading comprehension, the researcher was conducted a several stages. First, the researcher employed the first research instrument called *The Updated Vocabulary Level Test (TUVLT)* developed by (Paul Nation, 1983) and updated by (Webb et al., 2017) focusing solely on 2000 and 3000-word levels of vocabulary only. The VLT is regarded as the most widely used instrument for assessing L2 word knowledge (Read, 2000). It is used to measure the senior high students' vocabulary knowledge. The test consists of 60 questions in total, of which each level consists of 30 questions. The test was distributed through offline meeting by using paper-based. Additionally, to know the students' reading comprehension in expository text, the researcher used multiple-choice reading comprehension test, particularly in expository text. There are 20 questions and 2 text in analytical exposition text included in the test. The test is used to assess the students' reading proficiency in comprehending expository text, which covers eight aspects of reading comprehension, such as understanding the main idea, expression or idiom, making inference, grammatical feature, stated detail, unstated detail, supporting idea, and vocabulary in context.

Second, to check the consistency of both tests, the researcher conducted a reliability test by using Cronbach's Alpha. The test was conducted based in the data that have been collected in instruments reliability try-out to the 30 students outside the sample of this study. Below are the results of reliability test.

Table 1. VLT Reliability Test

Reliability Statistics	
Cronbach's Alpha	N of Items
.913	60

Table 2. Reading Comprehension Reliability Test

Reliability Statistics	
Cronbach's Alpha	N of Items
.827	20

According to the VLT reliability test by using Cronbach's Alpha, the level of reliability was .913. According to Cohen (2007), an instrument is considered reliable if its Cronbach's alpha value is 0.70 or higher. This makes the

questionnaire trustworthy. Additionally, the reliability test of reading comprehension test show that the level of reliability is .827. It suggests that the reading comprehension test was reliable to be administered in the next research, since the value is greater than >0.7 .

Third, the data collection process was carried out to the two selected classess with a total of 65 students as the research participants. The test was administered during English lesson hours for approximately two hours. Before starting the test, the researcher provided a clear instructions and explained the procedures to avoid any misunderstanding from the students. The vocabulary test was distributed first, followed by the reading comprehension test. Each test has been given a time allocation and the researcher monitored the students during the process to ensure that the test runs independently and in an orderly manner. After the completion of both tests, the students' answer sheets were collected and prepared for the scoring stage.

Moreover, after all the data were collected, there are several techniques in data analysis applied by the researcher. The researcher employed descriptive statistics analysis in the SPSS 27.0 program to address the first and second research questions. This type of analysis serves to present the data in a clearer and more concise format. After that, in answering the last research questions, the next data analysis used is Pearson product moment correlation. It is used to measure the level of correlation between variables. However, before calculating the correlation between vocabulary knowledge and reading comprehension of expository text, the normality test by Kolmogorov-Smirnov is conducted in order to know whether the score distribution is normal or not. The result indicated that the data was normally distributed, since the p value exceeds 0.05 (p value=0.2000 for both VLT and reading comprehension test).

Findings

There are several findings presented in this sub-section concerning the students' vocabulary knowledge and reading comprehension in expository text and its correlation.

Students' Vocabulary Knowledge

This section attempts to answer the first research question i.e. What is senior high students' vocabulary knowledge at various word levels. Once the data had been collected, the

researcher examined it through descriptive statistical analysis. Below is the result.

Table 3. Descriptive Statistics Results of VLT Test

	N	Minimum	Maximum	Mean	Std. Deviation
VLT Test	65	11	58	36.28	11.457
Valid N (listwise)	65				

Based on the result collected from The Updated Vocabulary Level Test score from the participants, there are 65 students (N=65) were in the moderate level of vocabulary knowledge at the 2000 and 3000-word level (Mean=36.28). This is strengthened by the result of standard

deviation (Std. Deviation=11.457), indicating a moderate to high degree of diversity in the students' vocabulary knowledge. The following categories apply to the vocabulary knowledge results of the students.

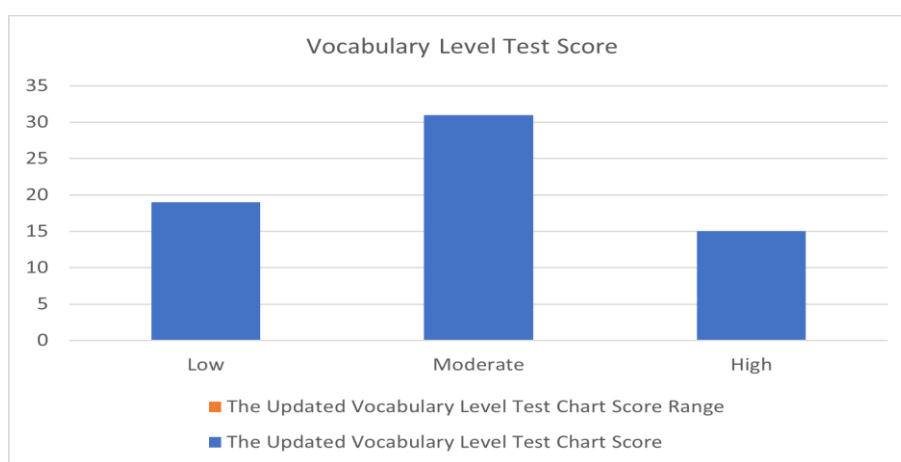


Figure 1. Classification of students' vocabulary knowledge

According to the chart above, the distribution of students' vocabulary scores can be observed. The findings show that the majority of students (31 students) possess a moderate level of vocabulary knowledge. This group is followed by those with low scores (19 students) and those with high scores (15 students), which represents the smallest proportion.

Students' Reading Comprehension in Expository Text

In order to answer the second research question, specifically, What level of reading comprehension do senior high school students possess when reading expository texts. The researcher administered a reading comprehension test focused on analytical exposition text to a sample of 65 students. The table below presents the results of the analysis of students' reading comprehension in expository texts, based on descriptive statistical analysis.

Table 4. Descriptive Statistics of Reading Comprehension Test

	N	Minimum	Maximum	Mean	Std. Deviation
Reading Comprehension Test	65	15	100	64.31	21.174
Valid N (listwise)	65				

According to the table above, the mean value of students' reading comprehension is 64.31. Indicating that, overall, students demonstrated a moderate capacity to comprehend expository texts. The scores varied between 15 as

the minimum and 100 as the maximum values obtained. The details of the reading comprehension obtained by students is present below.

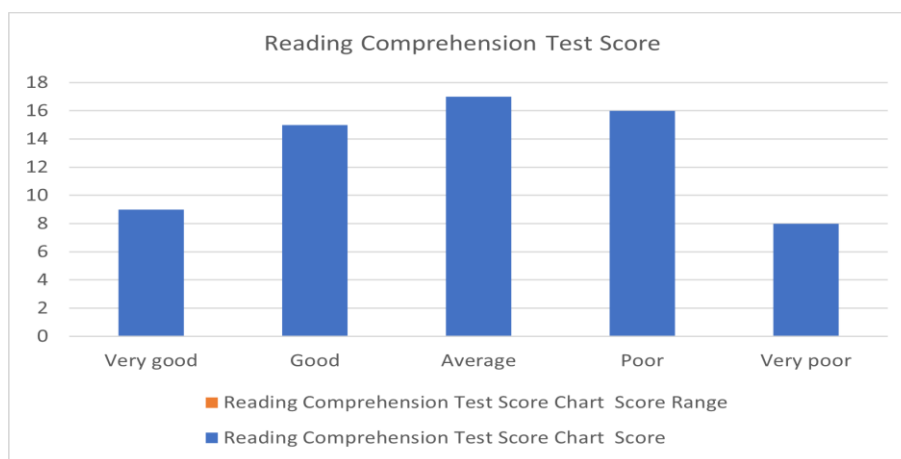


Figure 2. Classification of students' reading comprehension in expository text

Referring to the distribution illustrated in the chart, the 17 major students in this sample were categorized as being an average level. While other students are in the poor category (16 students), good category (15 students), very good category (9 students), very poor (8 students).

Correlation Between Vocabulary Knowledge and Reading Comprehension in Expository Text

The final data analysis conducted in this study addresses the third research question; Is there any correlation between senior high students' reading comprehension of expository text and their vocabulary knowledge at various word levels. To investigate the correlation between the variables, this study employed the Pearson Product-Moment Correlation. The results of the correlation analysis are presented below.

Table 6. Pearson Correlation Analysis Between Variables

		VLT Test	Reading Comprehension Test
VLT Test	Pearson Correlation	1	.792**
	Sig. (2-tailed)		<.001
	N	65	65
Reading Comprehension Test	Pearson Correlation	.792**	1
	Sig. (2-tailed)	<.001	
	N	65	65

As shown in the table, the VLT test and reading comprehension test had a correlation coefficient (*r* value) of .792. This is shown as a significant positive relationship between students' vocabulary knowledge and their comprehension of expository text. This means the coefficient correlation value shown in the table was classified as strong correlation. In addition, the *p*-value (Sig. 2-tailed), often known as the significance value, is less than 0.05, or < .001, indicating that the correlation is statistically significant at the 1% significance level.

Discussions

Based on the descriptive statistics analysis, the results show that students' vocabulary knowledge was at a moderate level, suggesting that neither students' was in low or high

vocabulary knowledge. Similar findings from (Royeras & Sumayo, 2024) has also reported that senior high students' vocabulary knowledge is generally at moderate level. This indicates that students at the basic to intermediate level of English vocabulary can identify general and common terms used in everyday conversations, and most students are able to understand familiar vocabulary at a basic to moderate level that they have previously encountered. Consequently, students possessing moderate to high levels of vocabulary knowledge generally achieve moderate success in reading comprehension tasks.

Furthermore, the researcher discovered that students' reading comprehension, particularly of expository texts, was found to be at a moderate level. This result indicates that

students' abilities remain at a moderate level as they continue to experience difficulties with reading comprehension. In other words, while some students demonstrated high levels of proficiency, others still exhibited lower levels of proficiency. Students at the intermediate level had sufficient vocabulary to understand the general meaning of texts. However, they struggled to manage and comprehend complex reading material.

Moreover, the results of this study reveal a notable positive relationship between students' vocabulary knowledge and their comprehension of expository texts. Therefore, based on the collected data, it can be said that students who perform well on the vocabulary level test are likely to achieve considerably better results on the reading comprehension test. The strong correlation observed in this study is supported by numerous studies whose results show some similarities in findings. First, (Masrai, 2019) found that vocabulary knowledge can predict how well students are able to comprehend reading, especially when dealing with informational texts such as expository which often require skills in decoding and academic language. Further findings from (Susanto, 2018) and (Manihuruk, 2020) which emphasized that having a strong vocabulary size can predict reading performance for high school students, especially in non-narrative texts where vocabulary recognition directly affects meaning-making. Additionally, (Ayu Cleverisa et al., 2022) specifically highlighted that students with lack of vocabulary will be more likely to face difficulties in understanding expository texts, which typically contain complex meanings and structures. As concluded by (Djuarsa, 2018) that students who have strong vocabulary knowledge at the mid frequency word level are more capable of handling the language requirements of expository texts. In addition, all of the above findings are in line with research by (Ma & Lin, 2015) which stated that the key to processing factual context efficiently is by increasing vocabulary knowledge. These findings reinforce the importance of vocabulary knowledge as a base element in comprehending expository text effectively.

CONCLUSION

According to the research objective, which aimed at examining the relationship between

students' vocabulary knowledge and their reading comprehension of expository texts, the researcher administered tests to 65 10th graders in total, at a state senior high school in Surabaya. Based on the research that have been conducted, the results indicated that students' vocabulary knowledge at the 2000-word level and 3000-word level were categorized as moderate level ($M = 36.28$). These findings suggest that certain students continue to show a limited grasp of vocabulary knowledge. At both levels, the majority of students struggled to identify the advanced vocabulary. Furthermore, in line with the earlier explanation, the students' reading comprehension results was categorized as average level with the mean score of 64.31. furthermore, the majority of students reported challenges in understanding various aspects of reading comprehension, including identifying detailed information, interpreting vocabulary in context, and determining the main idea. Particularly, based on the Pearson correlation coefficient value which indicate the r value of 0.792. This shows a strong positive relationship between students' vocabulary knowledge and their ability to comprehend expository text. This suggest that students with higher vocabulary knowledge generally demonstrate better reading comprehension skills, and vice versa. Thus, the findings of this study are intended to increase our understanding of the importance of integrating vocabulary knowledge with reading comprehension, particularly expository text, which are a common type of text that are taught in school curricula. Based on the results, there are several recommendations offered for further researchers regarding the result of this study. Since vocabulary knowledge is proved to have positive correlation on reading comprehension in expository text, further researchers may examine other variables, such as knowledge of grammar, use of reading strategies, or learners' motivation. While vocabulary knowledge has been demonstrated to be correlate with reading comprehension, but other factors might also influence the outcome. Moreover, to provide a more comprehensive understanding of how vocabulary impacts students' reading comprehension, other researchers could incorporate a qualitative approaches or instruments, including interviews and either open- or close-ended questionnaires. Therefore, these recommendations will yield more thorough and detailed results for any future research.

ACKNOWLEDGMENT

The researcher would like to express sincere gratitude to the lecturers for their constructive feedback and insightful suggestions which enabled the completion of this research. Moreover, appreciation is also conveyed to the teacher and all of the participants of this research, for the time, opportunity, and willingness to participate in this study.

REFERENCES

- Alshehri, A. S. (2024). *The relationship between reading comprehension and critical thinking among EFL learners*. 23(3). <https://doi.org/10.21831/ltr.v23i3.74118>
- Ayu Cleverisa, S., Sahiruddin, S., & Perdhani, W. C. (2022). The Correlation between EFL Students' Vocabulary Mastery and Their Reading Ability. *Jurnal Pendidikan : Riset Dan Konseptual*, 6(3), 487. https://doi.org/10.28926/riset_konseptual.v6i3.555
- Baker, M. (2018). *In Other Words*. Routledge. <https://doi.org/10.4324/9781315619187>
- Bombo, A., Gde, I., Agus Pramerta, P., Gede, D., & Kumara, A. G. (2022). Students' Vocabulary and Reading Comprehension: Is It Correlated? *Academic Journal on English Studies*, 2(2).
- Castillo et al., 2022; Elleman & Oslund, 2019; C. E. Snow & Uccelli. (2024). Exploring the impact of personalized and adaptive learning technologies on reading literacy: A global meta-analysis. *Educational Research Review*, 42.
- Creswell John W., C. J. D. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research (4th ed.)*.
- Dakhi, S. (2019). The Principles and the Teaching of English Vocabulary: A Review Tira Nur Fitria. *Journal of English Teaching*, 5(1). <https://ssrn.com/abstract=3378646>
- Djuarsa, H. M. (2018). *The Relationship Between Vocabulary Knowledge and Reading Comprehension Ability of Senior High School Students*.
- Etikan, I. (2016). Comparison of Convenience Sampling and Purposive Sampling. *American Journal of Theoretical and Applied Statistics*, 5(1), 1. <https://doi.org/10.11648/j.ajtas.20160501.11>
- Gallagher, M. A., Barber, A. T., Beck, J. S., & Buehl, M. M. (2019). Academic Vocabulary: Explicit and Incidental Instruction for Students of Diverse Language Backgrounds. *Reading and Writing Quarterly*, 35(2), 84–102. <https://doi.org/10.1080/10573569.2018.1510796>
- Ma, Y., & Lin, W. (2015). A Study on the Relationship between English Reading Comprehension and English Vocabulary Knowledge. *Education Research International*, 2015, 1–14. <https://doi.org/10.1155/2015/209154>
- Manihuruk, D. H. (2020). The Correlation between EFL Students' Vocabulary Knowledge and Reading Comprehension. *JET (Journal of English Teaching)*, 6(1), 86–95. <https://doi.org/10.33541/jet.v6i1.1264>
- Masrai, A. (2019). Vocabulary and Reading Comprehension Revisited: Evidence for High-, Mid-, and Low-Frequency Vocabulary Knowledge. *SAGE Open*, 9(2). <https://doi.org/10.1177/2158244019845182>
- Paul Nation. (1983). *Testing and Teaching Vocabulary*.
- Perfetti, C., & Stafura, J. (2014). Word Knowledge in a Theory of Reading Comprehension. *Scientific Studies of Reading*, 18(1), 22–37. <https://doi.org/10.1080/10888438.2013.827687>
- Read, J. A. S. (2000). *Assessing vocabulary*. Cambridge University Press.
- Royeras, J. T., & Sumayo, G. S. (2024). Vocabulary Knowledge and Inferential Reading Comprehension of Senior High School Students: A Descriptive-Correlational Inquiry. *East Asian Journal of Multidisciplinary Research*, 3(3), 1143–1154. <https://doi.org/10.55927/eajmr.v3i3.8164>
- Susanto, A. (2018). Vocabulary Knowledge in Relation to Student's Reading Comprehension: A Review. *INOVISI JOURNAL*, 3(1).
- Webb, S., Sasao, Y., & Ballance, O. (2017). The updated Vocabulary Levels Test. *ITL - International Journal of Applied Linguistics*, 168(1), 33–69. <https://doi.org/10.1075/itl.168.1.02web>