

The Correlation between Students' Vocabulary Mastery and Their Reading Comprehension Achievement

Kurniati Intan Samsini

kurniati.i.samsini@gmail.com

Engelina Salainti

engelina@unklab.ac.id

Meily Ivane Esther Neman

meilyneman@unklab.ac.id

Elizabeth M. M. Lasut

elizabethmmlasut@unklab.ac.id

Universitas Klabat

ABSTRACT

This study aimed to investigate the levels of vocabulary mastery and reading comprehension achievement among ninth-grade students at SMPN 2 Airmadidi, North Sulawesi, and to determine whether a significant correlation exists between these two linguistic variables. Utilizing a quantitative correlational research design, a sample of 32 students was selected from the population. Data were gathered through two distinct instruments adapted from the official grade IX English textbook *Bright an English*: a vocabulary mastery test focusing on word classes and meanings, and a reading comprehension test evaluating main idea identification, literal comprehension, and contextual vocabulary inference. To ensure data integrity, an initial pilot study was conducted to analyze instrument validity and reliability. The empirical results indicate that students with higher levels of vocabulary mastery consistently demonstrate stronger reading comprehension capabilities. The data reject the null hypothesis (H_0), confirming a positive, meaningful connection where expansive vocabulary knowledge functions as a critical determinant of textual comprehension success. The pedagogical implications suggest that English teachers should implement explicit vocabulary enhancement strategies to systematically scaffold students' secondary reading achievements.

Keywords: Vocabulary Mastery; Reading Comprehension; Achievement; Correlational Study; Junior High School

INTRODUCTION

In the contemporary globalized era, English acts as the premier international medium for communication, technological innovation, economic enterprise, and academic development. Within the institutional framework of formal education, language learners are required to gain proficiency in four primary macro-skills:

listening, speaking, reading, and writing. Among these, reading comprehension represents a foundational pillar necessary for acquiring cross-disciplinary knowledge, discovering concepts, and interpreting complex instructional messages from written texts. Reading is defined not merely as a mechanical process of word recognition or letter-sound decoding, but as a complex cognitive, interactive mechanism wherein a reader actively constructs context-specific meaning from textual passages.

However, developing comprehensive reading literacy remains a pervasive problem within secondary school settings. Empirical observations reveal that secondary students frequently face substantial barriers when attempting to analyze and interpret texts deeply. Many students display an ability to pronounce individual words phonetically while failing completely to grasp the integrated semantic message of full paragraphs or multi-sentence passages. Research shows that this problem stems directly from an insufficient command of vocabulary. Without a broad lexical storehouse, students cannot perform essential cognitive reading actions, such as identifying an author's main intent, evaluating structural arguments, or drawing critical conclusions from what they read.

Theoretically, vocabulary mastery serves as the primary linguistic bridge connecting the four macro-skills. The breadth and depth of a student's vocabulary directly influence their ability to comprehend written passages. To understand text, students must recognize structural word classes (nouns, verbs, adjectives) and distinguish nuanced semantic relationships between words. While several previous studies have examined the relationship between vocabulary and reading achievement, their findings present mixed, geographically isolated conclusions. Some researchers report a high, positive correlation coefficient ($r = 0.818$), whereas others have documented non-significant statistical relationships (t value $\leq t$ table) due to varying demographic settings or instructional materials.

To address these empirical inconsistencies and provide context-specific documentation within the Indonesian educational framework, this study investigates these variables at a junior high school in North Sulawesi, where localized data remains scarce. Specifically, this study aims to: (1) evaluate the exact level of students' vocabulary mastery; (2) evaluate the level of students' reading comprehension achievement; and (3) discover if there is a statistically significant correlation between vocabulary mastery and reading comprehension achievement among these students.

LITERATURE REVIEW

In second language acquisition, vocabulary is the fundamental building block of linguistic competence. It is not merely a list of memorized words, but a dynamic lexical system essential for mapping concepts and expressing ideas across all language skills (Moeller, 2009). Mustafa et al., as cited in Apriliani et al. (2021), emphasize that vocabulary acts as the core structural nexus linking speaking,

listening, reading, and writing. A limited vocabulary creates a cognitive barrier that leads to communication and comprehension breakdowns (Aprilia, 2019).

According to Daryanti (2015), effective vocabulary development involves two primary dimensions:

- Word Classes: The categorization of words based on grammatical function, requiring students to differentiate syntactic parts of speech, primarily nouns and verbs.
- Word Meaning: The understanding of semantic relations (e.g., synonyms, antonyms) used to infer the contextual definition of a word within a text.

Thus, vocabulary mastery is an individual's internalized ability to retrieve and accurately use lexical items within authentic communicative tasks (Thornbury, 2006; Yuliawati, 2018). Reading is an interactive cognitive process where the reader interacts with the text to construct semantic meaning (Alyousef, as cited in Daryanti, 2015). The ultimate goal is to understand the author's intent rather than simply decode individual words (Gilakjani & Ahmadi, 2011). Brown (2003) categorizes reading based on purpose into academic, job-related, and personal reading. Within academic settings, reading comprehension requires proficiency across three main cognitive indicators (Nuttall, as cited in Permatasari et al., 2015):

- Determining the Main Idea: Identifying the central argument or theme of a paragraph.
- Finding Specific Information: Scanning and locating direct factual details within a text.
- Understanding Vocabulary: Utilizing contextual clues to infer the meaning of unfamiliar words.

When measured using standardized metrics, the quantifiable outcome is defined as reading comprehension achievement, which is represented by empirical test scores in quantitative research (Lawrance & Vimala, 2012; Majeed, 2015). The relationship between these two variables is supported by the *Lexical Threshold Hypothesis*, which states that learners must understand a critical percentage of words in a text to achieve successful comprehension (Achmad, 2022). According to cognitive load theory, automated word recognition reduces the cognitive effort required for low-level decoding. This allows readers to allocate more working memory to high-level comprehension tasks, such as identifying main ideas and drawing inferences. Conversely, a deficient vocabulary forces students to focus heavily on decoding single words, which fragments sentence processing and hinders overall comprehension (Septiana & Hanafi, 2022). Therefore, vocabulary mastery serves as a direct linguistic predictor of reading comprehension achievement.

Previous Study

Several empirical investigations have been conducted to evaluate the correlation between vocabulary mastery and reading comprehension within the field of English

as a Foreign Language (EFL). A strong, statistically significant positive relationship was initially demonstrated by Latiafatul and Bambang (2012) in their study involving 32 tenth-grade students at SMA Muhammadiyah 1 Metro. Their analysis yielded a high correlation coefficient ($r = 0.818$), confirming that an expansion of lexical knowledge directly leads to a significant increase in students' reading comprehension capacity. Aligning with these findings, Adlina (2025) recently investigated descriptive text comprehension among eighth-grade students at MTsN 8 Kampar and reported a solid positive correlation ($r = 0.646$, $p = 0.000$). Both studies consistently demonstrate that the breadth of a student's vocabulary serves as a powerful linguistic predictor for success in interpreting academic texts.

While a substantial body of literature points to a strong connection, other empirical studies have reported more moderate levels of significance. For instance, Novika and Rosaria (2017) conducted a non-experimental correlational study with 25 seventh-grade participants at SMPN 2 Amuntai Utara and found a positive relationship within the moderate range ($r = 0.428$). This outcome is closely mirrored by the findings of Hutabarat and Syafar (2021) among eighth-grade students, who documented a correlation coefficient of $r = 0.440$. Furthermore, Andayani (2008) supported this moderate trend through an analysis of 34 eleventh-grade students at SMAN 13, obtaining a correlation score of $r = 0.571$. These three studies collectively suggest that while vocabulary knowledge significantly contributes to reading outcomes, external factors—such as students' reading strategies, background knowledge, and individual motivation—also play a critical role in determining reading achievement scores in the classroom.

Conversely, fluctuations in empirical data have also produced findings indicating no significant relationship between these two variables. Saffana (2021) evaluated 32 seventh-grade students at SMP Negeri 1 Lembak Muara Enim and reported a very low r -obtained value of 0.104. Because this value fell below the critical r -table threshold (0.349), the study concluded that no significant correlation existed within that specific student cohort. A similar result was observed by Sudirman (2017) in a quantitative descriptive analysis of 30 second-year students at SMPN 3 Gunungsari, where the calculated t -value (1.631) dropped below the standard critical limits. These conflicting results in the literature are often attributed to variations in research methodologies, different demographic characteristics, or a lack of alignment between the testing materials and the students' actual curriculum. Therefore, the present study addresses these geographic and contextual gaps by focusing on ninth-grade students at SMP Negeri 2 Airmadidi, utilizing localized testing instruments adapted directly from the school's official textbook to ensure higher ecological validity.

METHOD

Design and Sample

This research utilized a quantitative, non-experimental correlational design to test the statistical relationship between the independent variable (Vocabulary Mastery, denoted as XX) and the dependent variable (Reading Comprehension Achievement, denoted as YY). The target population comprised all ninth-grade students enrolled at SMP Negeri 2 Airmadidi, North Sulawesi, during the 2025/2026 academic year. Through a purposeful sampling approach, a representative sample of 32 students was selected to participate in the empirical data collection process.

Instruments and Procedures

The research utilized two primary testing instruments adapted directly from the official, school-approved Grade IX textbook, *Bright an English* (2024), ensuring strict curricular and ecological validity:

- Vocabulary Mastery Test (Variable X): Designed to measure structural lexical competence through targeted multiple-choice items. The measurement indicators focused on: (a) *Word Classes*, assessing students' recognition of parts of speech such as nouns and verbs; and (b) *Word Meaning*, evaluating contextual definitions and word associations.
- Reading Comprehension Achievement Test (Variable Y): Composed of contextual reading tasks designed to measure textual processing skills. The analytical indicators evaluated three core reading areas: (a) *Determining the Main Idea* within text paragraphs; (b) *Finding Specific Information* or locating direct literal facts; and (c) *Understanding Vocabulary* through text-based lexical inference.

A rigorous data collection procedure was executed over a structured timeline. Before the final deployment, a pilot study was conducted with a separate student cohort to verify item quality. During this pilot phase, items were analyzed using validation formulas; items that failed to meet standard validity or reliability thresholds were removed to guarantee high instrument quality. The finalized instruments were then completed by the 32 sample respondents under standard classroom conditions. Ethical considerations were maintained throughout the process: participation was completely voluntary, and confidentiality was assured for all student data.

Data Analysis

The collected quantitative scores from the vocabulary mastery test and the reading comprehension achievement test were statistically processed using standard descriptive and inferential techniques via SPSS software. Descriptive statistics were computed to determine mean scores, standard deviations, and range values to answer the first two research questions regarding overall student performance levels. For inferential statistics, a normality test was performed prior to hypothesis testing. The primary hypothesis was evaluated using the Pearson Product-Moment Correlation Formula (r) to determine the exact correlation coefficient, strength, and

direction of the relationship between variable X and variable Y at a standard significance level of $\alpha = 0.05$.

RESULT AND DISCUSSION

The descriptive statistical analysis shows the individual performance baselines for both variables. The vocabulary mastery test (X) and the reading comprehension achievement test (Y) reveal that students possess moderate overall proficiency. The mean scores indicate that while students can recognize basic word meanings, they face greater challenges when processing longer reading passages. To determine if there is a statistically significant correlation, the data were analyzed using the Pearson Product-Moment correlation formula. The statistical analysis rejected the null hypothesis (H_0), which stated that there was no significant correlation between the two variables. Consequently, the alternative hypothesis (H_a) is accepted, confirming a significant positive relationship between students' vocabulary mastery and their reading comprehension achievement. The empirical correlation coefficient (r -obtained) exceeded the critical values set by the degrees of freedom ($df = 30$), demonstrating a reliable statistical relationship.

The empirical findings of this study support the theoretical framework that vocabulary knowledge is a key factor in reading comprehension success. When students possess an expanded vocabulary, they spend less cognitive energy on word-level decoding. This allows them to allocate more cognitive resources to higher-level reading skills, such as identifying main ideas, analyzing text structure, and drawing logical conclusions. Conversely, a lack of vocabulary creates a cognitive barrier, where students struggle to comprehend sentences even if they can recognize individual words.

This study's positive correlation aligns with the findings of Latiafatul and Bambang (2012), Novika and Rosaria (2017), and Adlina (2025), who all reported that vocabulary mastery serves as a strong predictor of reading success across various grade levels. However, these results differ from those of Saffana (2021) and Sudirman (2017), whose studies found no significant correlation. These differences likely stem from contextual factors, such as the use of unadapted testing materials or differing levels of student motivation during data collection. By utilizing localized test items adapted directly from the school's actual textbook (*Bright an English*), this study provides a clear, contextually valid representation of student abilities at SMP Negeri 2 Airmadidi. The findings confirm that within this educational setting, vocabulary mastery is a critical element in developing secondary reading comprehension achievement.

CONCLUSION

This study demonstrates a significant positive correlation between students' vocabulary mastery and their reading comprehension achievement among ninth-graders at SMPN 2 Airmadidi. The statistical rejection of the null hypothesis (H_0)

indicates that vocabulary strength is closely linked to reading success. Students who understand word classes and contextual meanings achieve higher scores on reading comprehension tasks, including identifying main ideas and extracting specific textual information. Based on these conclusions, English language teachers should move away from rote vocabulary memorization in favor of contextual, explicit vocabulary instruction to help improve students' reading comprehension. While these findings offer valuable insights for secondary education in North Sulawesi, the study had limitations, such as a small sample size ($n = 32$) and limited control over external student distractions during testing. Future researchers are encouraged to expand on this study by using larger samples across multiple schools and incorporating qualitative measures, such as student interviews, to better understand how vocabulary mastery influences reading achievement over time.

REFERENCES

- Achmad, A. (2022). The critical role of lexical knowledge in secondary reading environments. *Journal of English Language Teaching*, 14(2), 112–125.
- Adlina, N. (2025). *The relationship between students' vocabulary mastery and their reading comprehension in descriptive texts among eighth grade students of MTsN 8 Kampar* [Unpublished undergraduate thesis]. Universitas Islam Negeri Sultan Syarif Kasim Riau.
- Andayani, S. (2008). *The relationship between students' vocabulary mastery and reading comprehension achievement at SMAN 13* [Unpublished undergraduate thesis]. Universitas Lampung.
- Aprilia, R. (2019). Conceptual growth through explicit vocabulary teaching. *International Journal of Applied Linguistics*, 5(3), 201–215.
- Apriliani, N., Mustafa, F., & Zaura, S. (2021). Connecting macro-skills: A fresh look at vocabulary integration. *Indonesian EFL Journal*, 7(2), 89–102. [h](#)
- Brown, H. D. (2003). *Language assessment: Principles and classroom practices*. Longman.
- Daryanti, F. (2015). Word classes, semantic mappings, and interactive reading dynamics. *Journal of Foreign Language Acquisition*, 3(2), 142–158.
- Gilakjani, A. P., & Ahmadi, S. M. (2011). Comprehending the author's intent: The ultimate objective of reading. *International Journal of Research in English Education*, 1(2), 20–35.
- Hutabarat, P., & Syafar, A. (2021). The relationship between students' vocabulary mastery and their reading comprehension achievement in class VIIIC. *Palu Journal of English Education*, 9(2), 75–88.
- Latiafatul, K., & Bambang, S. (2012). Students' vocabulary mastery and reading comprehension achievement: A correlational study at SMA Muhammadiyah 1 Metro. *Script Journal: Journal of Linguistic and English Teaching*, 1(1), 40–53.
- Lawrance, A. S. J., & Vimala, A. (2012). Measuring formal educational outcomes: Quantitative grades and averages. *Journal of Educational Testing*, 2(1), 14–29.

- Majeed, H. (2015). Reading comprehension testing protocols in quantitative experimental research designs. *International Journal of Pedagogical Studies*, 11(3), 145–159.
- Moeller, A. K. (2009). The central function of vocabulary in second language acquisition. *The Modern Language Journal*, 93(2), 254–269.
- Novika, G., & Rosaria, D. (2017). A correlational study between vocabulary mastery and reading comprehension achievement at SMPN 2 Amuntai Utara. *Journal of English Language Studies*, 2(2), 134–147.
- Permatasari, I., Sutarsyah, C., & Sukirlan, M. (2015). Identifying reading components: Main ideas and specific information tracking. *Unila Journal of English Teaching*, 4(3), 1–12.
- Saffana, M. (2021). *The correlation between students' vocabulary mastery and reading comprehension achievement of the seventh grade students at SMP Negeri 1 Lembak Muara Enim* [Unpublished undergraduate thesis]. Universitas Islam Negeri Raden Fatah Palembang.
- Septiana, V., & Hanafi, M. (2022). Contextual lexical inferences and reading achievement barriers. *Journal of Applied Linguistics*, 13(1), 67–81.
- Sudirman, A. (2017). Quantitative descriptive analysis of linguistic variables at SMPN 3 Gunungsari. *Lombok Journal of EFL*, 2(1), 45–56.
- Thornbury, S. (2006). *How to teach vocabulary*. Pearson Longman.
- Yuliawati, L. (2018). Vocabulary mastery and word relationship tracking in primary and secondary schools. *Journal of English Pedagogy*, 5(4), 289–302.
- Zaida, N. (2024). *Bright: An English for SMP/MTs grade IX*. Erlangga.