

Relationship between Vocabulary Knowledge and Inferential Reading Comprehension in Junior High Students

Tarisyia Amelia Indana Zulfa ¹⁾, Eva Rahmawati ²⁾

^{1,2} English Education, Faculty of Languages and Arts, Universitas Negeri Surabaya, Indonesia

Email Correspondence: evarahmawati@unesa.ac.id

ARTICLE INFO

Keywords:

[Vocabulary Knowledge;
Inferential Reading
Comprehension]

Article History:

Received : 21-07-2026

Revised : 13-08-2026

Accepted : 13-08-2026

Published : 20-08-2026

ABSTRACT

Vocabulary knowledge is critical component of reading comprehension, particularly inferential comprehension, which requires readers to interpret implicit meaning beyond explicitly stated information. Although previous studies have confirmed the importance of vocabulary for reading comprehension, limited research had examined how vocabulary knowledge across different word-frequency levels related to inferential comprehension among junior high school students. This study aimed to identify students' level of vocabulary knowledge, describe their inferential reading comprehension, and determine whether a significant relationship existed between the two. A quantitative correlational design was employed with 53 eighth-grade students from a private junior high school in Sidoarjo. Vocabulary knowledge was measured using the Updated Vocabulary Levels Test (1000-5000 levels), and inferential reading comprehension was assessed using an inference-based test adapted from TOEFL Junior materials. Data were analyzed using descriptive statistics and Spearman's rho. Results showed strong vocabulary mastery at the 1000-word level (92.5%), which declined and became more varied from 2000- through 5000-word levels. Inferential reading comprehension yielded a mean score of 57.83, falling into the sufficient category. A significant moderate positive correlation was found between vocabulary knowledge and inferential reading comprehension ($r = .488$, $p = .000$), strongest at the 1000-word ($r = .513$) and 2000-word ($r = .541$) levels. These findings indicated that vocabulary mastery at the basic and intermediate levels served as a primary predictor of inferential reading performance, suggesting that teachers should prioritize structured, frequency-informed vocabulary

instruction paired with explicit inference practice to support students' transitions toward higher-order reading skills.

1. Introduction

Reading comprehension has long been regarded as one of the central competencies in second and foreign language learning, given its direct contribution to learners' ability to acquire information, build knowledge, and participate meaningfully in academic life. For learners of English as a Foreign Language (EFL), reading frequently functions as the primary channel of language input, particularly in settings where opportunities for face-to-face interaction with native speakers are scarce but access to written English is comparatively abundant (Saeedi et al., 2016). Through sustained engagement with texts, learners are exposed not only to vocabulary and grammatical patterns but also to higher-order cognitive operations such as analyzing, synthesizing, and evaluating information (Castles et al., 2018). For this reason, reading comprehension is widely treated as the ultimate goal of reading instruction and is considered one of the strongest indicators of a student's broader academic trajectory.

Reading comprehension itself is not a monolithic skill. Building on Barrett's three-level taxonomy of reading comprehension (as cited in Brassell & Rasinski, 2008), it is commonly described as unfolding across three hierarchical levels: literal comprehension, in which the reader identifies information that is explicitly stated in a text, such as names, places, or events mentioned directly; inferential comprehension, in which the reader draws on background knowledge to uncover meanings and implications that are not directly expressed; and critical comprehension, the most demanding level, in which the reader evaluates a text's validity, quality, and relevance against personal experience and external standards. This three-level structure is echoed in Koda's (2005) account of the components underlying comprehension, which identifies decoding knowledge, vocabulary knowledge, and syntactic knowledge as the building blocks readers rely on to move from surface-level word recognition toward deeper meaning-making. Among the three levels of comprehension, inferential comprehension occupies a particularly important position, since it requires readers to merge textual cues with prior knowledge in order to reconstruct meanings the author has left implicit (Hall et al., 2020). This process draws on several distinct cognitive operations, including predictive inference, in which readers anticipate what will happen next (Graesser et al., 1994); bridging (or connective) inference, in which readers link separate pieces of explicit information into a coherent whole (Singer et al., 1992); elaborative inference, in which readers add detail or context drawn from their own experience (Clinton et al., 2020); and evaluative inference, the most demanding of the four, in which readers form judgments about an author's attitudes, motivations, or unstated message (Cooter et al., 1986). Together, these operations allow readers to build a coherent and complete mental representation of a text rather than a fragmented list of isolated facts (Kispaal, 2008; Noordman & Vonk, 2015). Crucially, each of these inferential operations is constrained by the quality of the reader's underlying lexical knowledge. The Lexical Quality Hypothesis proposed by Perfetti (2007) asserts that the high-quality representation in lexical aspects such as orthography,

phonology, semantics, and syntax allow for easier identification of words and also support higher level comprehension processes such as making inferences. This mechanism directly underlies the inferential operations described earlier. Readers can hold multiple explicit propositions active long enough to connect them into bridging inferences, can access the fuller network of semantic associations a word affords in order to supply elaborative detail, can anticipate upcoming content in predictive inference, and can register the subtler connotative and pragmatic cues that signal an author's attitude in evaluative inference. When lexical representations are low-quality or only partially specified, by contrast, word identification itself becomes effortful, leaving few resources available for the inferential comprehension but as a plausible cognitive mechanism through which vocabulary knowledge enables it.

Of the three levels, inferential comprehension is consistently identified as the most difficult EFL learners to master, largely because it demands more advanced cognitive processing and reasoning than literal comprehension (Hall et al., 2020). A learner may be entirely capable of locating explicit facts in a text yet still struggle to determine what those facts imply. In the Indonesian context, this issue carries particular weight, as the *Kurikulum Merdeka* explicitly foregrounds the development of deep textual understanding and the ability to infer implicit meaning as part of its English learning outcomes for junior high school students (Kementerian Pendidikan, Kebudayaan, Riset dan Teknologi, 2022). However, empirical evidence shows a significant gap between literal and inferential comprehension abilities. Previous research by Rizki Asani (2022) revealed that Indonesian students tend to succeed in literal comprehension tasks, most of them still struggle when the same text requires inferential reasoning, with up to 63% of students failing at the inferential level. This pattern indicates that the transition from literal to inferential comprehension is not an automatic byproduct of general reading ability, but instead depends on learners possessing sufficient linguistic and cognitive resources to support the inferential process.

Among the linguistic factors thought to underlie this transition, vocabulary knowledge has received sustained attention as one of the most influential (Shi, 2023). Vocabulary knowledge is not just about knowing the meaning of words. According to Nation (2022), this also involves understanding the form, meaning, and use of words in various situations. These aspects are generally described under two major dimensions which are vocabulary breadth, referring to the total amount of vocabulary knowledge a learner possesses, and vocabulary depth, referring to the degree of knowledge a learner has about a particular word, including semantics, relationships, and appropriate use (McGregor et al., 2013). This breadth-depth distinction reflects the same lexical-quality mechanism proposed by Perfetti (2007). Breadth determines how many word a reader can bring a well-specified representations to bear on, while depth determines how fully specified any single representation is across its orthographic, phonological, semantic, and syntactic components. Learners with well-developed vocabulary, in both breadth and depth, are therefore better positioned to connect the meanings of words to textual context and prior knowledge, improving their ability to draw inferences from what they read.

However, learning vocabulary is seen as a continuing challenge for EFL students in Indonesia. The *Kurikulum Merdeka* situates the expected English proficiency of junior high school students (Phase D) at the B1 level of the Common European Framework of Reference (CEFR) (Kementerian Pendidikan, Kebudayaan, Riset dan Teknologi, 2022). It is worth noting that the Learning Outcomes document itself defines the B1 target primarily in terms of communicative competencies (e.g., sustaining interaction, conveying main ideas, and making basic inferences from implicit information) rather than a fixed vocabulary count. Independent vocabulary-size research nonetheless offers an approximate benchmark for what a B1-level learner might be expected to know, with estimates in the literature ranging from around 2750 to over 3250-word families depending on the estimation method used (Milton & Alexiou, 2009). This range is used here only as an approximate point of reference and should not be read as an official numerical target embedded in the curriculum document. This goal has not been fully achieved by most students therefore with this limited vocabulary, students' understanding of implied meaning and reading comprehension of information requiring inferential comprehension is limited.

Consequently, vocabulary knowledge and inferential reading comprehension are closely interconnected. Studies by Binder et al. (2017), Ma & Lin (2015), and Gallagher et al. (2019) provide evidence that learners who have a wider and deeper vocabulary acquire and demonstrate greater reading comprehension. In addition, each study has shown that the contribution of vocabulary knowledge to the learning of inferential comprehension needs further research, particularly when applied to learning English in a foreign country, such as Indonesia.

A growing body of research has established that vocabulary knowledge is a strong predictor of reading comprehension in EFL contexts. Studies consistently demonstrate that both vocabulary breadth and depth are significantly associated with learners' reading performance, indicating that students who know more words and understand them more deeply tend to comprehend texts more successfully (Al Qunayeer, 2021; Alsowat, 2022; Dagnaw, 2023; Rabadi, 2023). In addition, vocabulary size has been found to correlate positively with learners' reading and writing achievement, further highlighting the central role of lexical knowledge in language development (Quines, 2023). These findings suggest that strengthening students' vocabulary knowledge is an effective approach to improving reading comprehension.

However, several limitations can be identified in the existing literature. Most previous studies have been conducted at the university level and international settings, with limited attention given to junior high school students, particularly in Indonesia. Moreover, reading comprehension has commonly been treated as a general construct, while specific sub-skills such as inferential reading comprehension remain underexplored. Critically, none of the studies reviewed above (Al Qunayeer, 2021; Alsowat, 2022; Dagnaw, 2023; Quines, 2023; Rabadi, 2023) examined vocabulary knowledge as a graded, multi-level construct in relation to inferential comprehension specifically. Vocabulary was instead treated as a single composite score correlated with general reading comprehension. There have also

been limited studies employing age-appropriate instruments that specifically measure inferential comprehension.

Therefore, this study intends to examine the relationship between vocabulary knowledge and inferential reading comprehension among junior high school students in Indonesia. Unlike prior research, which has generally correlated a single overall vocabulary score with general reading comprehension among university-level or adult EFL learners, this study distinguishes itself by (a) disaggregating vocabulary knowledge into five word-frequency levels (1000-5000) to examine whether the strength of its relationship with inferential comprehension varies across levels, (b) focusing specifically on inferential comprehension rather than reading comprehension as a whole, and (c) situating the investigation within the Indonesian junior high school context under the *Kurikulum Merdeka*, a population and policy setting that remains largely absent from existing literature. This research aims to (1) identify the level of vocabulary knowledge of junior high school students, (2) describe the performance of junior high school students in inferential reading comprehension, (3) determine whether there is a significant relationship between the level of vocabulary knowledge and inferential reading comprehension of junior high school students. The findings of this study are expected to contribute theoretically to a more specific understanding of how vocabulary knowledge, and its distribution across frequency levels, supports inferential reading and pedagogically to inform English teachers about the importance of integrating frequency-informed systematic vocabulary instruction to enhance students' higher-level reading comprehension.

2. Methods

This study employed a quantitative descriptive-correlational design to describe junior high school students' vocabulary knowledge and inferential reading comprehension, and to examine the relationship between the two variables without any manipulation (Ary et al., 2009). The descriptive component was intended to portray students' performance on vocabulary knowledge and inferential reading comprehension, while the correlational component was employed to determine whether a statistically significant relationship existed between the two variables.

The population of this study consisted of all eighth-grade students of a private junior high school in Sidoarjo, distributed across seven classes (three International Class Orientation/ICO classes, one Tahfidz class, and three Excellent classes), totaling 201 students. The ICO classes were prioritized over the Tahfidz and Excellent classes because, within this school's tracking system, ICO students receive a comparatively greater number of weekly English contact hours and are taught using English-medium instructional materials, making them the subgroup most likely to have developed the vocabulary and reading skills targeted by this study. A purposive sampling technique was applied to select participants considered representative of the research objectives (Ary et al., 2009). Since the three ICO classes follow the same intensive English curriculum, they were considered relatively homogenous in terms of English learning experience. Two ICO classes (8A and 8C) were selected based on the English teacher's recommendation, considering relatively

balanced academic performance and intensive exposure to English, as well as class schedule feasibility, since 8A and 8C were the two classes whose schedules aligned with the timeframe available for data collection, so that the remaining ICO class was not included for this reason. The final sample comprised 53 eighth-grade students, which exceeds the minimum sample size recommended for correlational research (Cohen et al., 2007). This sample size, drawn from a relatively homogenous subpopulation and exceeding the recommend minimum, further supports the adequacy of the sample for the statistical aims of this study.

Two instruments were used for data collection, such as the Updated Vocabulary Levels Test (VLT) by Webb et al. (2017) and Inferential Reading Comprehension Test adapted from The Official Guide to the TOEFL Junior® Standard Test (ETS, 2014) and Master TOEFL Junior Intermediate (CEFR Level B1): Reading Comprehension by Hahn (2014). The VLT measures receptive vocabulary knowledge across five frequency levels (1000–5000-word families) using a matching format. Each level consisted of 10 clusters of three definitions and six words (three keys and three distractors). There were 30 definitions at each level. The test was distributed to students through Google Forms.

The Inferential Reading Comprehension Test was adapted from the two sources through the following procedure. From The Official Guide to the TOEFL Junior® Standard Test (ETS, 2014), the researcher screened the 42 multiple-choice items in the reading section and retained only items classified as inference questions, since most items measured literal comprehension, vocabulary, or reference skills rather than inference. This screening yielded five inference items drawn from three reading passages (two items from the first passage, two from the second, and one from the third). To supplement this limited number, 15 additional items were adopted from Unit 6: Inference Questions in Master TOEFL Junior intermediate (CEFR Level B1): Reading Comprehension Hahn (2014). This unit is specifically designed to assess inferential reading, so all items were already inference-based and did not require further selection. Combining the two sources produced a 20-item instrument that pairs standardized TOEFL Junior items with a sufficient number of items specifically targeting inference.

The 20 items were classified according to four established typed of inference. Table 1 presents the blueprint of the Inferential Reading Comprehension instrument, including the descriptions of each inference type and its corresponding item numbers

Table 1. Blueprint of the Inferential Reading Comprehension Instrument

Inference Type	Description	Item Numbers	N
Bridging	Connecting two or more explicitly stated details within the text to draw a local conclusion (Singer et al., 1992)	1, 2, 3, 4, 5, 8, 10, 13, 14, 17, 20	11
Elaborative	Drawing on background knowledge or schema to build a situation model that goes beyond information directly linked in the text (Clinton et al., 2020)	6, 7, 9, 12, 15, 18	6
Evaluative	Judging characters or authors' motivations, emotions, attitudes, or reputations that are not explicitly stated (Cooter et al., 1986)	11, 16, 19	3

Predictive	Anticipating forthcoming events based on cues from an unfolding, sequential narrative (Graesser et al., 1994). Not represented in this instrument (see note below).	–	0
Total			20

As shown in Table 1, none of the 20 items specifically targeted predictive inference. Predictive inference relies on cues from an unfolding, sequential narrative that allow readers to anticipate forthcoming events, a text structure that is largely absent from the short, self-contained expository and narrative passages characteristic of the TOEFL Junior reading format used as the basis for this instrument. In addition, the inferential reading test consisted of 20 multiple-choice items and distributed to students in a paper-based format.

2.1 Validity Test

Valid means that the instrument is suitable or can be used to measure what should be measured (Ary et al., 2009). A valid instrument is one of the requirements to get a valid result of research (Sugiyono, 2021). Both instruments were reviewed by an expert validator, namely a lecturer in the English Education study program at State University of Surabaya, with expertise in language testing and assessment, and was therefore qualified to judge the content relevance and linguistic appropriateness of both instruments for junior high school test-takers. the validity of both instruments was measured using a validation sheet with a five-point Likert scale (1 = poor to 5 = excellent). The validation sheet covered nine indicators grouped under six broader aspects, such as clarity (of individual items and instructions), content accuracy (appropriateness of the item format for measuring the intended construct), relevance (alignment of items with the research objectives), content validity (accuracy of the information the items convey), and language accuracy (comprehensibility and correctness of the language used). The Updated VLT obtained a validity score of 34 out of a maximum of 45, while the inferential reading comprehension test obtained a score of 35 out of 45. Based on these results, both instruments were judged suitable for use without revision, thereby establishing their content validity.

2.2 Reliability Test

Reliability refers to the score test consistency or the stability from the instrument whether stable and consistent or not (Ary et al., 2009). The reliability of both instruments was estimated using Cronbach's Alpha, with a coefficient of .70 regarded as an acceptable threshold for internal consistency (Taber, 2018). As presented in Table 2, all VLT levels and the inferential reading comprehension test yielded Cronbach's Alpha coefficients above .70, indicating that both instruments were reliable for use in this study.

Table 2. Reliability of the Instruments

Instrument	N of Items	Cronbach's Alpha
VLT Level 1000	10	.925
VLT Level 2000	10	.895
VLT Level 3000	10	.888

VLT Level 4000	10	.913
VLT Level 5000	10	.903
Inferential Reading Comprehension Test	20	.807

The collected data were analyzed using SPSS version 26. Both instruments were conducted offline in the classroom during regular English lessons on two separate occasions under the supervision of the researcher and the classroom teacher.

2.3 Descriptive Statistics

Descriptive statistics (minimum, maximum, mean, and standard deviation) were used to summarize students' scores on both variables.

Students' VLT score were analyzed using the raw scores obtained, with each correct answer given a score of 1 and a maximum score of 30 at each level. In addition, students' vocabulary mastery was descriptively classified based on the cut-off scores proposed by Webb et al. (2017), as presented in Table 3. Students were considered to have mastered the 1000, 2000, and 3000 levels if they scored between 29 and 30, whereas mastery of the 4000 and 5000 levels was determined by a score between 24 and 30. This classification was used only to describe the extent of students' vocabulary mastery at each level and did not alter the numerical scores used in the subsequent statistical analysis.

Table 3. Criteria for Vocabulary Level Mastery Based on the Vocabulary Levels Test

Vocabulary Level	Score Range	Category
1000 Level	29-30	Mastered
	0-28	Not Mastered
2000 Level	29-30	Mastered
	0-28	Not Mastered
3000 Level	29-30	Mastered
	0-28	Not Mastered
4000 Level	24-30	Mastered
	0-23	Not Mastered
5000 Level	24-30	Mastered
	0-23	Not Mastered

For the inferential reading comprehension test, each correct answer was given a score of 1 and each incorrect answer a score of 0. To facilitate data interpretation, the raw scores obtained by the students are converted into a 100-point scale using the following formula:

$$Final\ Score = \frac{\sum Raw\ Score}{20} \times 100$$

To determine students' ability level, the mean score was consulted against the five achievement categories proposed by Arikunto (2018), as shown in Table 4.

Table 4. Classification of Students' Inferential Reading Ability Levels

Score Range	Criteria
0 – 20	Very Bad

21 – 40	Bad
41 – 60	Sufficient
61 – 80	Good
81 – 100	Very Good

Prior to correlation analysis, normality and linearity tests were conducted to determine the appropriate statistical technique.

2.4 Normality Test

Normality test was examined using the Kolmogorov-Smirnov test, which is appropriate for a sample size greater than 50 (Mishra et al., 2019). Data are considered normally distributed when the significance value (p) is greater than 0.05 (Pallant, 2020). As shown in Table 5, all variables yielded significance values below 0.05, indicating that the data were not normally distributed. Consequently, a non-parametric statistical technique was required for the correlation analysis.

Table 5. Normality Test Results

	Kolmogorov-Smirnov		
	Statistic	df	Sig.
VLT_1000	.253	53	.000
VLT_2000	.233	53	.000
VLT_3000	.123	53	.043
VLT_4000	.249	53	.000
VLT_5000	.215	53	.000
VLT Total Score	.205	53	.000
Inferential Reading Score	.169	53	.001

a. Lilliefors Significance Correction

2.5 Linearity Test

A formal linearity test was not required for the correlation analysis in this study, as Spearman's rho assesses monotonic rather than strictly linear association (Field, 2018). Nevertheless, a supplementary Test of Linearity through one-way ANOVA was conducted to provide additional descriptive information on the shape of the relationship between each VLT subtest and inferential reading comprehension scores. The linearity assumption is met when the significance value of the Linearity term is below 0.05 and the significance value of Deviation from linearity is above 0.05 (Ghozali, 2018). As shown in Table 6, most VLT levels (1000, 2000, 3000, 4000, 5000, and the total score) showed a significant linearity term with a non-significant deviation from linearity, indicating a predominantly linear pattern, while VLT 3000 showed a significant deviation from linearity ($p = .005$), suggesting a less linear relationship at this level. As Spearman's rho does not require linearity, this finding does not affect the validity of the correlation analysis reported in this study.

Table 6. Linearity Test Results

	Sig. Linear Term	Sig. Deviation from Linearity	Linearity Pattern
VLT 1000	.000	.938	Linear
VLT 2000	.000	.377	Linear
VLT 3000	.000	.005	Not Linear
VLT 4000	.011	.247	Linear
VLT 5000	.007	.111	Linear
VLT Total Score	.001	.295	Linear

2.6 Hypothesis Testing

Since the data were not normally distributed, non-parametric correlation analysis was employed using Spearman's rho. Data are classified as significantly correlated if the correlation coefficient is less than 0.05. on the other hand, the data can be classified as not significantly correlated if the coefficient correlation is greater than 0.05. The strength of correlation was interpreted using the criteria proposed by Cohen et al. (2007), as shown in Table 7.

Table 7. Correlation coefficient interpretation

Correlation Value Ranging	Degree of Correlation Coefficient
0.20 – 0.35	Low
0.35 – 0.65	Moderate
0.65 – 0.85	High
>0.85	Very High

The hypothesis tested in this study are formulated as follows:

1. H_0 (Null Hypothesis): There is no significant relationship between vocabulary knowledge and inferential reading comprehension.
2. H_1 (Alternative Hypothesis): There is a significant relationship between vocabulary knowledge and inferential reading comprehension.

If the significance value (p) obtained from Spearman's rho is less than 0.05, H_0 is rejected and H_1 is accepted. In contrast, if p is greater than 0.05, H_0 is accepted and H_1 is rejected.

3. Result

3.1 Students' Vocabulary Knowledge

This section attempts to answer the first research question i.e. What is the level of vocabulary knowledge of junior high school students. Once the data had been collected, the researcher examined it through descriptive statistical analysis to investigate the mean, minimum, maximum, and standard deviation of students' VLT score. The results are presented in Table 8.

Table 8. Descriptive Statistics of Vocabulary Levels Test

Vocabulary Level	N	Min	Max	Mean	SD
1000 Level	53	18	30	27.68	3.030
2000 Level	53	5	30	23.89	6.393
3000 Level	53	5	30	20.08	8.211
4000 Level	53	1	30	20.34	8.665
5000 Level	53	2	30	21.53	7.934

As shown in Table 8, the highest vocabulary mastery is found at Vocabulary Level 1000 with the mean score of 27.68 and a standard deviation of 3.030. At vocabulary level 2000, the mean score decreased to 23.89 (SD = 6.393). The mean scores for Vocabulary Levels 3000, 4000, and 5000 were recorded between 20.08 and 21.53, with a standard deviation consistently above 7, indicating that students' scores at these levels were more widely spread than the previous levels.

Based on the mastery criteria proposed by Webb et al. (2017), students are classified into the Mastered and Not Mastered categories using different cut-off scores at each level (see Table 3). The distribution of students' vocabulary mastery classification is presented in Table 9.

Table 9. Distribution of Students' Vocabulary Mastery

Vocabulary Level	Category	Frequency (n)	Percentage (%)
1000 Level	Mastered	49	92.5
	Not Mastered	4	7.5
	Total	53	100
2000 Level	Mastered	33	62.3
	Not Mastered	20	37.7
	Total	53	100
3000 Level	Mastered	22	41.5
	Not Mastered	31	58.5
	Total	53	100
4000 Level	Mastered	31	58.5
	Not Mastered	22	41.5
	Total	53	100
5000 Level	Mastered	30	56.6
	Not Mastered	23	43.4
	Total	53	100

As shown in Table 9, 92.5% of students (n = 49) were classified as Mastered at the 1000-word level. This proportion decreased to 62.3% (n = 33) at the 2000-word level and further to 41.5% (n = 22) at the 3000-word level. At the 4000- and 5000-word levels, the proportion of students classified as Mastered rose again to 58.5% (n = 31) and 56.6% (n = 30), respectively. This apparent increase reflects the different mastery cut-off scores applied at each level (a score of 29-30 was required at the 1000-3000 levels, compared with 24-30 at the 4000-5000 levels) rather than an actual improvement in students' vocabulary knowledge.

3.2 Students Inferential Reading Comprehension

In order to answer the second research question, specifically, how is the performance of junior high school students in inferential reading comprehension. The Inferential Reading Comprehension Test was administered to 53 students, and the results were analyzed descriptively, as presented in Table 10.

Table 10. Descriptive Statistics of Inferential Reading Comprehension.

	N	Minimum	Maximum	Mean	Std. deviation
Inferential Reading Comprehension Test	53	15	95	57.83	22.415
Valid N (listwise)	53				

Based on the table above, the minimum score that the students got in the inferential reading test is 15 and the maximum score is 95, with a mean score of 57.83 (SD = 22.415).

Students' inferential reading ability levels are grouped into five achievement categories based on Arikunto (2018) (see Table 4). The frequency distribution of students' ability categories is presented in Table 11.

Table 11. Inferential Reading Ability Categories.

Category	Frequency	Percent	Valid Percent	Cumulative Percent
Very Bad	2	3.8	3.8	3.8
Bad	13	24.5	24.5	28.3
Sufficient	17	32.1	32.1	60.4
Good	9	17.0	17.0	77.4
Very Good	12	22.6	22.6	100.0
Total	53	100.0	100.0	

The data in Table 11 show that the most frequently achieved category by students is the Sufficient category, with a total of 17 students (32.1%). Furthermore, there are 13 students (24.5%) in the Bad category, followed by 12 students (22.6%) in the Very Good category. The remaining 9 students (17.0%) are in the Good category, and the smallest number, 2 students (3.8%) are in the Very Bad category.

3.3 The Relationship between Vocabulary Knowledge and Inferential Reading Comprehension

The final data analysis conducted in this study addresses the third research question; Is there a significant relationship between the level of vocabulary knowledge and inferential reading comprehension of junior high school students. To investigate the correlation between variables, this study employed the Spearman's rho correlation. The results of the correlation analysis are presented below.

Table 12. Correlation between Variables

		1000	2000	3000	4000	5000	Total VLT
Inferential Reading Comprehension	Spearman's rho	.513**	.541**	.424**	.323*	.349*	.488**
	Sig. (2-tailed)	.000	.000	.002	.018	.010	.000
	N	53	53	53	53	53	53
95% CI		[.271, .694]	[.306, .711]	[.135, .654]	[.059, .565]	[.077, .595]	[.248, .692]
**. Correlation is significant at the 0.01 level (2-tailed).							

As shown in Table 12, the correlation between the Total VLT Score and inferential reading comprehension was $r = .488$, 95% CI [.248, .692], $p = .000$. This shows a significant positive relationship between students' vocabulary knowledge and their inferential reading comprehension, classified as a moderate correlation. A more detailed analysis of each vocabulary level revealed varying degrees of correlation strength. At the 1000, 2000, and 3000 levels, the results showed a moderate positive correlation with correlation coefficients of $r = .513$, 95% CI [.271, .694], $p = .000$; $r = .541$, 95% CI [.306, .711], $p = .000$; and $r = .424$, 95% CI [.135, .654], $p = .002$, respectively. These three levels are categorized as a moderate degree of correlation. In contrast, the correlation results at the 4000 and 5000 levels showed a slightly different pattern. At the 4000 level, a positive correlation was found with a coefficient of $r = .323$, 95% CI [.059, .565], $p = .018$, and at the 5000 level, the coefficient was $r = .349$, 95% CI [.077, .595], $p = .010$. The correlation results of these two levels are categorized as a low degree of correlation. Notably, the confidence intervals for the 4000- and 5000-word levels are also the widest and extend closest to zero, indicating less precise estimates than those obtained for the 1000- and 2000-word levels. However, both levels still remain statistically significant, as their p -values are lower than the 0.05 threshold. In short, it can be concluded that all levels of vocabulary knowledge have a significant positive relationship with inferential reading comprehension, with the overall strength of the relationship predominantly falling into the moderate category.

Thus, the alternative hypothesis (H_1) stating that there is a significant relationship between vocabulary knowledge and students' inferential reading ability is accepted, while the null hypothesis (H_0) is rejected.

4. Discussion

4.1 Students' Vocabulary Knowledge

An interesting pattern emerges when the Mastered/Not Mastered classifications are examined across word frequency levels: mastery rates decline steadily from the 1000- to the 3000-word level but then appear to rise again at the 4000- and 5000-word levels. This apparent improvement, however, should not be read as evidence that students' actual vocabulary knowledge increased at these higher, less frequent word levels. Rather, it is largely an artifact of the different cut-off scores used across levels, following Webb et al.

(2017). A stricter cut-off of 29-30 out of 30 was applied at the 1000-3000 levels, while a more lenient cut-off of 24-30 out of 30 was applied at the 4000-5000 levels. A clearer picture emerges when the raw mean scores in Table 8 are considered instead of the Mastered/Not Mastered classification. Students' performance in fact continued to decline gradually across the higher levels, from a mean of 20.08 (SD = 8.211) at the 3000-word level to 20.34 (SD = 8.665) at the 4000-word level and 21.53 (SD = 7.934) at the 5000-word level, with means and standard deviations remaining broadly similar across the three levels. This suggests that students' vocabulary knowledge did not genuinely improve at the 4000- and 5000-word levels. Instead, a larger proportion of students simply crossed the lower threshold set for those levels. This distinction is important to avoid misinterpreting the higher Mastered percentages at the 4000- and 5000-word levels as evidence that students command rarer, lower-frequency vocabulary more successfully than they do the 3000-word level.

Evidence from the declining score reinforces the premise that vocabulary knowledge is a complex process because it involves a multifaced understanding of word forms and their usage in various different contexts (Nation, 2022). The low-test scores observed at levels 4000 and 5000 indicate that although students may have limited vocabulary breadth or the number of words they know, they are still significantly lacking in the aspect of vocabulary depth, which includes a deep understanding of semantic relationships and the appropriate use of words (McGregor et al., 2013). The condition where students find it difficult to achieve the mastered category at 3000 to 5000 levels reflects the principle that the lower the frequency of a word, the greater the demand on students to understand the depth of its meaning in order to achieve full mastery.

This phenomenon is supported with the findings of Alsowat (2022), which showed that students' vocabulary scores tend to decrease as the frequency level of words becomes more less common (from level 1000 to higher levels). Many students in the study also were still lacking in the 95% vocabulary mastery threshold required to understand texts independently. This finding is similar to the data from this study, where students' proficiency at the 3000 and above are still below the optimal category. Furthermore, *Kurikulum Merdeka* stipulates that junior high school students (Phase D) are targeted to reach CEFR level B1, a competency benchmark that independent vocabulary-size research (rather than the curriculum document itself) associates with roughly 2750 to over 3250-word families depending on the estimation used (Milton & Alexiou, 2009). Using the lower end of that range as an approximate reference point, the present results show that only 62.3% of students mastered 2000-word level (41.5% at the stricter 3000-word criterion). This suggests that the curriculum's communicative B1 target has not been fully achieved, and students' vocabulary knowledge is still developing toward the expected standard for junior high school level.

4.2 Students' Inferential Reading Comprehension

The mean inferential reading comprehension score of 57.83%, which fall into the Sufficient category, is comparable to the findings of Al Qunayeer (2021), who noted that

intermediate-level EFL learners generally demonstrate moderate performance in reading comprehension. However, since this specific focuses on inferential comprehension, it provides a clearer overview of why a large proportion of students (28.3%) still fall into the bad and very bad categories. This outcome is also consistent with the research by Alsowat (2022), which identified that many EFL students remain below mastery threshold required to comprehend texts independently. Interestingly, the average performance in the current study shows a slightly better result than the poor category in that previous research.

Furthermore, these results underscore the importance of readers' prior knowledge in constructing meaning from a text. As inferential comprehension requires students to integrate textual clues with their background knowledge to derive implied meanings, the moderate scores suggest that many students may face difficulties in activating their stored knowledge during the reading process (Brassell & Rasinski, 2008). This aligns with the principle that reading is an interactive process where understanding is achieved not only by recognizing words but also by establishing causal relationships and predicting outcomes based on what the reader already knows (Kispal, 2008; Noordman & Vonk, 2015). The sufficient achievement of the students indicates that while they can recognize some implicit details, they still need more linguistic and cognitive resources to consistently achieve the deeper understanding required by the *Kurikulum Merdeka*.

4.3 The Relationship between Vocabulary Knowledge and Inferential Reading Comprehension

The results show a positive, moderate relationship between vocabulary knowledge and inferential reading comprehension, indicating that students with higher vocabulary knowledge scores also tend to have higher inferential reading scores, while those with lower vocabulary knowledge scores tend to show lower inferential reading scores. Since this study employed a correlational design, this pattern indicates an association between the two variables rather than a demonstrated causal effect. It should not be read as evidence that vocabulary knowledge directly produces or determines inferential reading ability. Establishing the direction and mechanism of any causal influence would require further regression-based or experimental research.

Theoretically, this pattern of association is consistent with the Lexical Quality Hypothesis (Perfetti, 2007), which proposes that higher-quality lexical representations are associated with more efficient higher-level comprehension processes, including inference-making. When students hold a stronger vocabulary, they may be able to connect word meanings with textual context more efficiently, which is consistent with a greater capacity to read between the lines. This interpretation aligns with Al Qunayeer (2021) and Quines (2023), who similarly observed that vocabulary mastery at the basic-to-intermediate level tends to accompany stronger reading performance, plausibly because vocabulary provides access to the linguistic cues that support the recognition of abstract or implicit concepts.

A closer look at the level-specific correlation coefficients and their confidence intervals (Table 12) helps explain why vocabulary knowledge at the 1000- and 2000-word levels was more strongly associated with inferential reading comprehension than knowledge at the 3000- to 5000-word levels. This pattern can be interpreted through a lexical threshold perspective, which proposes that learners must first attain a sufficiently high level of mastery over the most frequent, high-utility vocabulary before that knowledge can meaningfully support higher-order comprehension processes such as inference-making (cf. Milton & Alexiou, 2009; Nation, 2022). Because high-frequency words occur far more often in running text and carry a disproportionately large share of the lexical coverage needed to understand a passage (Nation, 2022), strong mastery of the 1000- and 2000-word levels most directly reduces the burden of lexical decoding, potentially freeing cognitive resources for the bridging, elaborative, and evaluative inferential operations (Graesser et al., 1994; Perfetti, 2007; Singer et al., 1992). In this sample, students had already achieved consistently high and comparatively homogenous mastery of the 1000-word level (92.5% Mastered, $SD = 3.03$), which may have allowed this foundational vocabulary knowledge to function as a threshold that differentiates weaker from stronger inferential readers. By contrast, vocabulary at the 3000- to 5000-word levels is encountered far less frequently in everyday input and school-based instruction, so students' knowledge of it was not only lower on average but also considerably more uneven, as reflected in the larger standard deviations reported in Table 8 ($SD > 7$ at the 3000-5000 levels compared with $SD = 3.03$ at the 1000-word level). Because these lower-frequency words are less central to constructing the overall meaning of a passage, individual differences in students' mastery of them appear to matter comparatively less for inferential comprehension once the foundational, high-frequency threshold has been met. This interpretation aligns with the theoretical position that lexical quality at the level of frequently encountered words most directly underlies the inferential operations readers rely on to construct meaning beyond what is explicitly stated (Perfetti, 2007).

The results of this study are broadly similar to the report by Rabadi (2023), which found a moderate correlation among EFL students, in which reading comprehension performance was closely associated with vocabulary performance. However, there is a difference in the strength of the relationship compared to the study by Dagnaw (2023), which reported a much stronger correlation of 0.73. The difference in distribution scores is most likely due to the use of different test instruments, considering the current study specifically focuses on inferential reading, which requires more complex reasoning than general reading. Thus, these findings therefore reinforce the argument that vocabulary mastery is an important correlate of students' ability to interpret texts deeply, consistent with current literacy demands. Because the present design is correlational, however, this association should not be interpreted as evidence that vocabulary mastery directly causes stronger inferential comprehension.

5. Conclusion

According to the research objective, which aimed to investigate students' level of vocabulary knowledge, their performance in inferential reading comprehension, and the relationship between the two, the researcher administered tests to 53 8th graders at a private junior high school in Sidoarjo. This conclusion focuses on the theoretical contribution, practical implications, and directions for future research that follow from them.

Theoretically, the level-specific patterns of correlation found in this study extends the Lexical Quality Hypothesis (Perfetti, 2007) by suggesting that the strength of the vocabulary-inference relationship is not uniform across word-frequency levels; the association was strongest at the 1000- and 2000-word levels and weaker, though still significant, at the 3000-5000-word levels. This graded pattern supports a lexical-threshold interpretation, in which mastery of high-frequency vocabulary functions as a foundational resource that frees cognitive capacity for bridging, elaborative, and evaluative inference-making, while knowledge of lower-frequency words plays a comparatively smaller role once this threshold is met. This contribution also distinguishes the present study from previous correlational research, which has generally treated vocabulary as a single composite score and reading comprehension as an undifferentiated construct, most often among university-level or adults EFL learners. By disaggregating vocabulary knowledge into five frequency levels, isolating inferential comprehension specifically, and situating the investigation within the Indonesian junior high school context under the *Kurikulum Merdeka*, this study extends the existing literature into a population, construct, and level of analytic granularity that earlier correlational studies have not addressed.

From a practical perspective, the level-specific pattern of association suggests that English instruction at the junior high school level should prioritize consolidating students' mastery of high frequency words (1000-2000) vocabulary before extensive attention is given to lower-frequency words, since it is this foundational vocabulary that appears most closely tied to students' capacity to draw inferences. Teachers are encouraged to pair vocabulary instruction with explicit inference practice, for example through guided questioning and contextual-clue analysis embedded directly in reading activities, rather than treating vocabulary and reading-strategy instruction as separate strands.

Because this study employed a correlational design, these findings describe an association rather than a causal mechanism, and this limitation points toward directions for future research. Subsequent studies could employ regression-based or experimental design to test whether targeted vocabulary instruction produces measurable gains in inferential reading comprehension, incorporate direct measures of vocabulary depth alongside the breadth-based VLT used here, and extend the investigation to larger and more diverse samples, including students outside intensive English tracks, to examine whether the lexical-threshold pattern observed here holds more broadly.

5. Acknowledgments

We would like to acknowledge to the lectures 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, who were willing to invest their time and energy to participate in this study.

References

- Al Qunayeer, H. S. (2021). An Investigation of the Relationship between Reading Comprehension, Vocabulary Knowledge, and English Language Proficiency Level of Saudi EFL Learners. *Advances in Language and Literary Studies*, 12(2), 59. <https://doi.org/10.7575/aiac.all.s.v.12n.2.p.59>
- Alsowat, H. H. (2022). The Mediating Roles of Gender and Academic Field in the Relationship between Vocabulary Knowledge and Reading Comprehension of EFL Students. *Universal Journal of Educational Research*, 10(2), 146–159. <https://doi.org/10.13189/ujer.2022.100204>
- Arikunto, S. (2018). *Dasar-Dasar Evaluasi Pendidikan* (3rd ed.). Bumi Aksara.
- Ary, D., Jacobs, L. C., & Sorensen, C. (2009). *Introduction to Research in Education* (8th ed.). Wadsworth Cengage Learning.
- Binder, K. S., Cote, N. G., Lee, C., Bessette, E., & Vu, H. (2017). Beyond breadth: the contributions of vocabulary depth to reading comprehension among skilled readers. *Journal of Research in Reading*, 40(3), 333–343. <https://doi.org/10.1111/1467-9817.12069>
- Brassell, Danny., & Rasinski, T. V. . (2008). *Comprehension that works : taking students beyond ordinary understanding to deep comprehension*. Shell Education.
- Castles, A., Rastle, K., & Nation, K. (2018). Ending the Reading Wars: Reading Acquisition From Novice to Expert. *Psychological Science in the Public Interest*, 19(2), 93. <https://doi.org/10.1177/1529100618786959>
- Clinton, V., Taylor, T., Bajpayee, S., Davison, M. L., Carlson, S. E., & Seipel, B. (2020). Inferential comprehension differences between narrative and expository texts: A systematic review and meta-analysis. *Reading and Writing* 33. <https://doi.org/10.1007/s11145-020>
- Cohen, L., Manion, L., & Morrison, K. (2007). *Research Methods in Education* (6, Ed.; 6th ed.). Routledge.
- Cooter, R. B., Joseph, D. G., & Flynt, E. S. (1986). Eliminating the Literal Pursuit of Reading Comprehension. *Journal of Clinical Raeding: Research and Practice*, 2.
- Dagnaw, A. T. (2023). Revisiting the role of breadth and depth of vocabulary knowledge in reading comprehension. *Cogent Education*, 10(1). <https://doi.org/10.1080/2331186X.2023.2217345>

- ETS. (2014). *Official Guide to the TOEFL Junior® Standard Test* (1st ed.). Educational Testing Service.
- Field, Andy. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). SAGE Publications.
- 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>
- Ghozali, Imam. (2018). *Aplikasi analisis multivariate dengan program SPSS* (9th ed.). Badan Penerbit Universitas Diponegoro.
- Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing Inferences During Narrative Text Comprehension. In *Psychological Review* (Vol. 101, Number 3).
- Hahn, R. (2014). *Master TOEFL Junior Intermediate (CEFR Level B1): Reading Comprehension*. Nhân Trí Việt.
- Hall, C., Vaughn, S., Barnes, M. A., Stewart, A. A., Austin, C. R., & Roberts, G. (2020). The Effects of Inference Instruction on the Reading Comprehension of English Learners With Reading Comprehension Difficulties. *Remedial and Special Education*, 41(5), 259–270. <https://doi.org/10.1177/0741932518824983>
- Kementerian Pendidikan, K. R. dan T. R. I. (2022). *Capaian Pembelajaran Mata Pelajaran Bahasa Inggris Fase A-Fase F untuk SD/MI/Program Paket A, SMP/MTs/Program Paket B, dan SMA/MA/SMK/MAK/Program Paket C*.
- Kispal, Anne. (2008). *Effective teaching of inference skills for reading : literature review*. Department for Children, Schools and Families.
- Koda, K. (2005). *Insights into Second Language Reading*. Cambridge University Press.
<https://doi.org/10.1017/CBO9781139524841>
- 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>
- McGregor, K. K., Oleson, J., Bahnsen, A., & Duff, D. (2013). Children with developmental language impairment have vocabulary deficits characterized by limited breadth and depth. *International Journal of Language and Communication Disorders*, 48(3), 307–319. <https://doi.org/10.1111/1460-6984.12008>
- Milton, J., & Alexiou, T. (2009). *Vocabulary Studies in First and Second Language Acquisition* (B. Richards, M. H. Daller, D. D. Malvern, P. Meara, J. Milton, & J. Treffers-Daller, Eds.). Palgrave Macmillan UK. <https://doi.org/10.1057/9780230242258>

- Mishra, P., Pandey, C. M., Singh, U., Gupta, A., Sahu, C., & Keshri, A. (2019). Descriptive statistics and normality tests for statistical data. *Annals of Cardiac Anaesthesia*, 22(1), 67–72. https://doi.org/10.4103/aca.ACA_157_18
- Nation, I. S. P. (2022). *Learning Vocabulary in Another Language* (3rd ed.). Cambridge University Press. <https://doi.org/10.1017/9781009093873>
- Noordman, L. G. M., & Vonk, W. (2015). Inferences in Discourse, Psychology of. In *International Encyclopedia of the Social & Behavioral Sciences: Second Edition* (pp. 37–44). Elsevier Inc. <https://doi.org/10.1016/B978-0-08-097086-8.57012-3>
- Pallant, J. (2020). *SPSS Survival Manual*. Routledge. <https://doi.org/10.4324/9781003117452>
- Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, 11(4), 357–383. <https://doi.org/10.1080/10888430701530730>
- Quines, Z. M. (2023). Impact of Students' Vocabulary Level to their Reading and Writing Performance. *International Journal of English Language and Linguistics Research*, 11(2), 18–32. <https://doi.org/10.37745/ijellr.13/vol11n21832>
- Rabadi, R. I. (2023). Examining the Role of Breadth and Depth of Vocabulary Knowledge in Reading Comprehension of English Language Learners. *Jordan Journal of Modern Languages and Literatures*, 15(1), 327–345. <https://doi.org/10.47012/jjml.15.1.17>
- Rizki Asani, F. (2022). *Students' Difficulties Analysis in Reading Comprehension at The Second Grade of MTs Nurul Ikhlas*.
- Saeedi, G., Khanmohammadi, A., & Mahmmodi, H. (2016). Reading Comprehension and Its Importance in Learning English as a Second Language. *International Journal of EFL*, 1(1), 61. <https://doi.org/10.21462/ijefl.v1i1.1>
- Shi, Y. (2023). How Inference Making Mediates the Relation Between Vocabulary Knowledge and Reading Comprehension for Chinese EFL Middle School Students. *English Teaching(South Korea)*, 78(3), 85–104. <https://doi.org/10.15858/engtea.78.3.202309.85>
- Singer, M., Andrusiak, P., Reisdorf, P., & Black, N. L. (1992). Individual differences in bridging inference processes. In *Memory & Cognition* (Vol. 20, Number 5).
- Sugiyono. (2021). *Metode penelitian kuantitatif, kualitatif, dan R&D* (3rd ed.). Alfabeta, CV.
- Taber, K. S. (2018). The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. *Research in Science Education*, 48(6), 1273–1296. <https://doi.org/10.1007/s11165-016-9602-2>

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>

Webb, S., Sasao, Y., & Ballance, O. (2017). The updated Vocabulary Levels Test: Developing and validating two new forms of the VLT. *ITL - International Journal of Applied Linguistics*, 168(1), 34–70.