

# Vocabulary Coverage and Reading Comprehension of University EFL Learners

Boniesta Melani\*, Sudirman Willian, Kurniawan Apagrianto, Husnul Lail

Universitas Mataram

\*Corresponding author. Email: [boniestamelani@unram.ac.id](mailto:boniestamelani@unram.ac.id)

## ABSTRACT

The knowledge of vocabulary is positively associated with second language learners' reading ability. Knowing at least 95% of vocabulary contained in the reading passages is necessary for adequate reading comprehension. This study measures Second Language (L2) learners' vocabulary coverage of six reading samples used as intensive reading materials for university EFL learners in Indonesia. Fifty second-year students enrolled in a Reading class were asked to mark difficult words or unknown words that they did not know the meaning and were unsure about. The type-token analysis of familiar and unfamiliar vocabulary undertaken through the *V\_Words* program showed that the learners knew around 98% of the total words (tokens) used in the texts, with averagely six to twelve words (dominated with nouns), were reportedly unknown in the 470 to 516-word length of texts. The 98% vocabulary coverage was estimated to assist the learners in sufficiently comprehending the reading materials without the need to check the meaning of unfamiliar words in dictionaries. Furthermore, having around 98% coverage of the vocabulary used in the texts provided more opportunity for the readers to guess the meaning of unfamiliar words and more flexibility for incidental vocabulary learning to take place during reading.

**Keywords:** EFL learners, Reading comprehension, Vocabulary coverage

## 1. INTRODUCTION

The knowledge of vocabulary has been shown to be positively associated with a second language (L2) learners' ability to read (Staehr, 2008). It is believed that vocabulary knowledge is a prerequisite for reading comprehension. The more vocabulary learners know, the more likely they understand a reading passage. It is also true that word knowledge is crucial for reading comprehension. Thus, a reader who knows only a few words tends to be a poor reader (Freebody & Anderson, 1983; Dixon, LaVevre, & Twilley, 1988; Nagy, Anderson, & Herman, 1987; Stahl et al., 1989). In other words, if a person's vocabulary knowledge is not limited to the number of words required, it does not support sufficient understanding. However, if a learner has crossed the limits of the required knowledge, understanding is possible for all learners (Nation & Hunston, 2013).

Having good understanding of reading materials is important, especially in university level, in which EFL learners have to deal with large number of English textbooks. Most of the books and scientific materials used as references in university are written in English. In

line with these findings, the Curriculum for English Departments of Teacher Training Faculties appointed by the Department of Education and Culture (1991) also discovered a large amount of English written references.

Studies focusing on the relationship of word knowledge and reading comprehension have concluded that word knowledge is influential for comprehension ability in reading (Dixon et al., 1988; Rashidi & Khosravi, 2010; Schmitt, Jiang, & Grabe, 2011). The breadth of vocabulary knowledge accounts for more than 60% of the variance in reading comprehension (Laufer, 1992; Moinzadeh & Moslehpour, 2012; Qian, 1999).

In this study, vocabulary knowledge refers to the knowledge of English words, which include knowing the forms, meaning and use of words (Nation & Hunston, 2013). Reading on the other hand is known as the process of making meaning from written texts (Anderson et al., 1985). Thus, reading comprehension refers to making meaning as a result of readers' interactions with reading text involving the combination of their prior knowledge and experience, information in the text, and their views of the text.

The amount of vocabulary needed for reading comprehension is an important aspect to see. A preliminary study on learners' ability to comprehend English for academic purposes concluded that learners need to gain reading scores at least around 65 to 70% for their reading comprehension to be called sufficient (Laufer & Sim, 1985). In a follow-up study to determine the percentage of words needed to be known to ensure comprehension, Laufer (1989) discovered that learners with 95% or more familiarity of the words appeared in the reading texts can significantly facilitate the learners to be successful readers (scored 55% or higher on reading test). The estimation was obtained by asking respondents to mark words in the text they did not know, who then assigned the numbers according to their responses on the word translation test.

In this study, the measurement of 90% did not show a significant difference. Having 95% coverage means that there are five unfamiliar words in every 100 tokens (see Table 1) or one unfamiliar word in two lines (Laufer & Sim, 1985).

**Table 1.** The number of unfamiliar words and text coverage

Coverage (%)	Unfamiliar words/ 100 Tokens	Lines / 1 Unfamiliar word
99	1	10
98	2	5
95	5	2
90	10	1
80	20	0.5

In addition, Schmitt et al. (2011) suggest that a 98% estimate is a more reasonable coverage target for readers of academic texts. Similarly, Hirsh and Nation (1992) suggest looking at around 98-99% of the vocabulary (or one unknown word in the text's 50-100 words) for reading pleasure or extensive reading activity. In Indonesia, there have not been many current studies in investigating the amount of vocabulary known by EFL readers in their reading materials or the learners' vocabulary coverage, and their ability to understand reading materials, especially in university level.

The vocabulary coverage needed by learners for learning vocabulary through reading differs between intensive reading and extensive reading. In intensive reading, which usually uses short reading texts (300-500 words), the purpose of reading is to understand the content of the reading, where the procedures applied include direct attention to vocabulary, grammar, and other language elements contained in the text. On the other hand, extensive reading focuses more on the meaning of words in reading and does not involve a lot of language use. Therefore, for vocabulary learning, based on the study of Paribakht and Wesche (1993), extensive reading texts should not contain more than 5% unfamiliar words (except nouns), and preferably no more

than 2% to ensure text understanding and the use of guessing strategies. In addition, texts that include the repetition of unfamiliar words provide more favorable conditions. To increase language proficiency, learners need to read text with little or no unfamiliar words at all. The appearance of unfamiliar vocabulary can slow down the reading process, making it difficult for learners to maintain reading flow. The following table shows the balance between the learners' vocabulary coverage and the type of reading as presented in Nation and Hunston (2013) (see Table 2).

**Table 2.** Types of reading and vocabulary coverage (Nation & Hunston, 2013)

Types of reading	Learning goals	Coverage
Intensive reading	Developing language Developing strategy use knowledge	Less than 95%
Extensive reading for vocabulary learning	Incidental vocabulary learning Reading skills	95-98%
Extensive reading for vocabulary mastery	Reading quickly	99-100%

Concerning vocabulary knowledge, studies conducted at different places in Indonesia show that high school graduates are likely to have a lesser amount of vocabulary size than the number of words expected to be learned (Melani et al., 2013; Nation & Hunston, 2013). The National curriculum requires high school students to learn approximately 4000 words in senior high school, 1500 words of which should have been learned in junior high school. Thus, English as a Foreign Language (EFL) learners should have a vocabulary size of around 4000–5000 English words when entering college. In a diagnosis of Indonesian EFL learners' reading problems, Nation and Hunston (2013) estimated that Indonesian university students' average English vocabulary recognition appeared in reading passages to be around 600 words. The students show unfamiliarity with basic vocabulary items such as pronouns, days of the week, numbers and high-frequently used verbs.

Looking at the time when the data were taken, it is predicted that the condition regarding the learners' knowledge of vocabulary has changed and shown more improvement than what has been experienced in the past. Recent research is needed to gain current information about the learners' vocabulary knowledge pertaining to their reading ability in university level. This study aims to estimate EFL learners' vocabulary coverage on the reading materials use at the university level in Indonesia. Specifically, it calculates university EFL learners' familiarity with word types and tokens found in the reading materials to estimate the learners' coverage of

English words in the texts for reading comprehension purposes. Furthermore, the study looks at the types of words that are likely to be unfamiliar to the learners.

## 2. METHOD

Fifty fourth-semester students (19 males and 31 females) around the age of nineteen to twenty-five-year-old participated in an intensive reading class. All of them are EFL learners with at least seven years of experience learning English as an L2 in a foreign language setting. They were asked to identify unknown vocabulary found in the reading materials used in an intensive reading class. Six reading materials of general topics (i.e., *Money*, *Canning Food*, *Pottery*, *Bringing Up Children*, *Auction Sale*, and *Social Classes*) used in the class were selected randomly from the materials used in the beginning, in the middle, and at the end of the semester. The materials were taken from a book on reading course entitled "Practical Faster Reading: An Intermediate/Advanced Course in Reading and Vocabulary" by Mosback and Mosback (1976). They are written for readers with an intermediate level of proficiency in English.

The calculation of Type/Token counts was conducted to check the lexical diversity of the reading materials. *Types* (different words) and *tokens* (running words), counts measure vocabulary diversity of a passage by calculating the number of different words that occur in it (Meara, 1978). This calculation was done by utilizing the *V\_Words* program, a small utility program that produces type and token counts for long texts (Meara & Miralpeix, 2016). This program was selected for its simple operation, and its sufficient function to do the type and tokens calculations of a text. Moreover, the program comes with a default of the GSL 1000 (K1), GSL 2000 (K2), and AWL, the features needed to profile the reading texts. It also provides word calculation result based on word types, families, and tokens. It is suggested to be used to analyse the written materials that the students are exposed to (Meara & Miralpeix, 2016). The program is available freely through <https://www.lognostics.co.uk/tools/index.htm>.

The students' vocabulary coverage is measured by calculating the percentage of known vocabulary in the texts (Laufer & Sim, 1985). The coverage estimation was obtained by asking respondents to firstly mark all

unfamiliar words that they found in the text. The results of this identification were later confirmed in a translation task, that is asking the students to translate the words that they did not know. The words that show partial knowledge were considered as known. To estimate the learners' text coverage, the calculation method used is as follows:

1. Number of known words = Total words in text - (number of unknown words + number of misinformed words x (number of differences x 100/40)).

2. Vocabulary coverage = number of words known in the text x 100 / number of words in the text.

The data gathered from the textual analysis, and the learners' word familiarity were later used to describe the learners' vocabulary coverage and predict their ability to cope with the comprehension of the reading materials.

## 3. FINDINGS AND DISCUSSION

The data gathered from the type-token analysis show that the length of the reading passages (i.e., *Money*, *Canning Food*, *Pottery*, *Bringing Up Children*, *Auction Sale*, and *Social Classes*) ranged from 470 to 516 running words or tokens, or 241 to 275-word types, with an average length of 493-word tokens (see Table 3). The results suggest that the texts used as the reading materials are of the same length. Looking at the length of the texts, these readings are suitable for intensive reading activity. In intensive reading, which usually uses short reading texts (e.g., 300-500 words), the purpose of reading is to understand the content of the reading. The activities involved in intensive reading may include direct attention to vocabulary, grammar, and other language elements contained in the text (Nation & Hunston, 2013). On the other hand, the text intended for extensive reading is usually moderate length, averagely 15-30 pages long. The text is long enough for readers to develop an idea or narrative fully and require them to keep track of ideas, concepts, or characters while reading. Thus, extensive reading focuses more on the meaning of words in reading and does not involve a lot of language use (Nation & Hunston, 2013). As reviewed in Cahyono and Widiawati (2006), EFL reading practice in Indonesia concentrates mostly on intensive reading, which involves close and careful reading for the purpose of achieving a study goal. Further analysis of the number of unfamiliar words identified by the learners in each of the texts shows that

**Table 3.** Learners' vocabulary coverage

Texts	Word tokens	Word types	Average unfamiliar words	Coverage (%)
Money	476	244	<b>6.36</b>	98.7
Canning Food	482	261	<b>9.64</b>	98
Pottery	511	262	<b>8.12</b>	98.4
Bringing Up Children	470	241	<b>6.8</b>	98.6
Auction Sale	516	275	<b>8.73</b>	98.3
Social Classes	502	269	<b>11.51</b>	97.7

there are around 6.4 to 11.5 words unknown by the students. This brings average unknown words in all of the six texts is around 8.5 words. The number of unknown words in one text is not far different from that of the other text, indicating that the texts have a similar level of difficulty. Those results are predicted since the texts used as reading materials in this class are intended for intermediate to advanced second language learners of English. Referring to the English proficiency level of the students in the program, which was based on their GPA level, the fourth-semester university EFL learners in this program are expected to be intermediate learners, and therefore, may have enough vocabulary knowledge for the given passages.

A calculation of the percentages of unfamiliar words and the total number of tokens in each text shows that the learners are familiar with approximately 97.7% to 98.7% (see Table 3) of the total words in all texts. According to these numbers the learners' vocabulary coverage based on the token's calculation was estimated to range from 97.7% to 98.7%. No previous studies to date specifically investigated the Indonesian university EFL learners' vocabulary coverage of the reading materials used in their reading course. Most studies focus on the learners' knowledge of vocabulary in general, or the size of vocabulary that the learners seem to be acquired (Melani et al., 2013; Nation & Hunston, 2013; Nurweni & Read, 1999; Quin, 1968). The results indicate that the learners know most of the English words used in the texts, meaning that they will find the materials easy to comprehend. Knowing 98% of the words in a text allows a second language reader to read independently without checking the meaning of unknown words in a dictionary. Moreover, Laufer (1989) predicted that for adequate reading comprehension, to score more than 55% in reading comprehension test, a reader needs to know more than 95% of the total vocabulary used in a passage. In addition, Hirsh and Nation (1992) suggest having a coverage of at least 98%, in which one word is unknown in every fifty to one hundred words, for independent reading. Given the percentage of familiar vocabulary that they have, the learners are predicted to sufficiently and independently comprehend all reading passages. Thus, knowing averagely 98% of the words in the texts gives no difficulty for the readers in this study to understand the reading materials.

Based on Paribakht and Wesche's (1993) study, extensive reading texts should not contain more than 5% foreign tokens (except nouns) and preferably no more than 2% to ensure text comprehension and use of guessing strategies. In this study, because the respondents indicated that the text's vocabulary coverage was around 98%, it could be estimated that the number of difficult words or unknown words was around 2% of the total words in the text. This coverage shows the high probability of language learners experiencing reading for pleasure in incidental (unintentional) conditions. No

direct intention to learn specific vocabulary in the text is required for the students to understand the content of the reading, both intentionally and incidentally (accidentally). With 98% coverage, it can be ascertained that the students can use a strategy to guess the meaning of unknown words by using the knowledge of familiar words, the contextual clues available surrounding the unknown words, and their general knowledge (Haastrup, 2008; Jenkins & Dixon, 1983). Studies focusing on incidental vocabulary learning from reading showed that small proportion of vocabulary could be learned accidentally through reading (Brown, Waring, & Donkaewbua, 2008; Gitsaki & Melani, 2013; Kweon & Kim, 2008; Teng, 2018). The learning gain, however, could be increased when the learner reads more comprehensive reading passages.

However, the students' high coverage of the reading materials may not be a good sign of having the texts for intensive reading activities. The vocabulary coverage needed by learners for learning vocabulary through reading differs between intensive reading and extensive reading. For intensive reading activities Nation and Hunston (2013), suggest that learners cover less than 95% of the vocabulary used in a text to be able to extend the development of their language proficiency and to develop the use of learning strategies. Thus, knowing more than 95% of the words in the texts does not give the learners the privilege to make the most of the reading materials to develop their language and strategic competencies. In this sense, EFL learners in university level may require upper level of reading materials that can challenge them to experience and apply more strategies to develop their reading skills, and thus, English language proficiency.

Figure 1 depicts the proportion of unfamiliar words in the texts and Table 4 exhibits the learners' vocabulary coverage. With regard to the types of unfamiliar words, the study found that difficult words were reported to be in some word categories (i.e., nouns, adjectives, adverbs, and verbs). Of all the unfamiliar vocabulary identified, most of the words reported to be unknown were nouns, followed by adjectives, verbs, and adverbs, respectively.

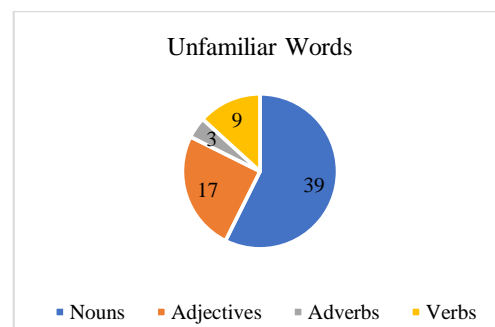


Figure 1 The proportion of unfamiliar words found in the texts.

**Table 4.** Learners' vocabulary coverage

Texts	Categories			
	Nouns ( <i>N</i> =39)	Adjectives ( <i>N</i> =17)	Adverbs ( <i>N</i> =3)	Verbs ( <i>N</i> =9)
Money	copper (K2) vessels (K1) cowries funerals (K2)	distinct (AWL) starve durable		hoard
Canning Food	proximity yeasts blanching wholesale dispatch decays (K2) tins (K2) wholesale	immersed enormous (AWL) trimmed	hermetically	corked (K2) deteriorate blanching
Pottery	kiln conjecture treadle pivot insulators platform pores utensils bonfire crafts	versatile scraping (K2) glazed porous		spin (K2)
Bringing Up Children	backwardness defects	interdependent tactile innate distinguished (K1)	reliably (AWL)	ascertain vary (AWL)
Auction Sale	auctioneer bidder hides (K2) spear rostrum spoils (K2) knock-out (K2)		extremely (K2)	assembled (AWL)
Social Classes	acquaintance peasant plebs fidelity predecessors burghers metrics remnants	recurrent vassal medieval		flourished

The fact that nouns are mostly unfamiliar is not surprising since nouns cover most of the words in the texts. From the 39 unknown nouns, 31 of which are above the 2000-word frequency level, seven are in the second 1000 GSL, and one is in the first 1000 GSL. This suggests that more than 80% of the total unknown nouns are not frequently used. Furthermore, almost three quarters of the total 68 unknown words (74.67%), are outside the 2000-word level of the General Service List GSL (West, 1953) and the Academic Word List AWL (Coxhead, 2000). Therefore, they are categorized as low-frequency words. In other words, almost one-fourth of the total unknown words are high-frequency words. This finding is relatively high, considering that the first 2000-word families are expected to have been acquired at high school level. Thus, the National curriculum requires high school graduates to have been known 4000–5000-word

families in English when entering college (Depdikbud, 1991).

#### 4. CONCLUSION

The necessity to know at least 95% of the total words in texts is pivotal for second language learners' ability to sufficiently comprehend reading passages written in a second language. The results of this study show that after having at least seven years of EFL instruction, second-year university learners view their familiarity with more than 95% of the vocabulary used in intensive reading materials. Of all the unfamiliar words, nouns are most likely unknown by the learners due to the high use of nouns in the texts. Knowing most of the words used in the texts (98% coverage) provides more opportunity for the learners to read the passages without dictionary

lookup independently. Thus, they can apply learning strategies, such as guessing the meaning of unknown words from context. Not only that the learners' vocabulary coverage is influential when the learners are engaged in reading activities, but it can also support the learning of unfamiliar words while reading. However, despite the need to have high vocabulary coverage for independent reading comprehension, less than 95% vocabulary coverage provides a preferable condition for intensive reading activities that facilitate the development of L2 learners' language and strategic competencies. It is expected that this finding provides an essential recommendation to the selection of intensive reading materials at the university level. For future research, more studies need to keep track of Indonesian EFL learners' vocabulary mastery to evaluate school instruction.

## REFERENCES

- Anderson, R. C., Hiebert, E. H., Scott, J. A., & Wilkinson, I. A. G. (1985). *Becoming a nation of readers*. Washington, D. C: National Institute of Education.
- Brown, R., Waring, R., & Donkaewbua, S. (2008). Incidental vocabulary acquisition from reading, reading-while-listening, and listening to stories. *Reading in a Foreign Language*, 20(2), 136-163.
- Cahyono, B. Y., & Widiati, U. (2006). The teaching of EFL reading in the Indonesian context: The state of the art. *Teflin Journal*, 17(1), 36-58. doi: 10.15639/teflinjournal.v17i1/37-60
- Depdikbud. (1991). *Kurikulum pendidikan tenaga kependidikan sekolah menengah program S1: Program studi pendidikan bahasa Inggris*. Jakarta: Author.
- Dixon, P., LeFevre, J. A., & Twilley, L. C. (1988). Word knowledge and working memory as predictors of reading skill. *Journal of Educational Psychology*, 80, 465-472. doi: 10.1037/0022-0663.80.4.465
- Freebody, P., & Anderson, R. C. (1983). Effects of vocabulary difficulty, text cohesion, and schema availability on reading comprehension. *Reading Research Quarterly*, 18, 277-294. doi: 10.2307/747389
- Gitsaki, C., & Melani, B. Z. (2013). 3 factors affecting incidental L2 vocabulary acquisition through authentic text reading: A study of Indonesian EFL learners. *The European Journal of Applied Linguistics and TEFL*, 2(1), 47-69.
- Haastrup, K. (2008). Lexical inferencing procedures in two languages. In D. Albrechtsen, K. Haastrup & B. Henriksen (Eds.), *Vocabulary and writing in a first and second language* (pp. 67-111). London: Palgrave Macmillan.
- Hirsh, D., & Nation, P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language*, 8(2), 689-696.
- Jenkins, J. R., & Dixon, R. (1983). Vocabulary learning. *Contemporary Educational Psychology*, 8(3), 237-260. doi: 10.1016/0361-476X(83)90016-4
- Kweon, S. O., & Kim, H. R. (2008). Beyond raw frequency: Incidental vocabulary acquisition in extensive reading. *Reading in a Foreign Language*, 20(2), 191-215.
- Laufer, B. (1992). How much lexis is necessary for reading comprehension? In P. J. Arnaud & H. Béjoint, (Eds.), *Vocabulary and applied linguistics* (pp. 126-132). doi: 10.1007/978-1-349-12396-4\_12
- Laufer, B., & Sim, D. D. (1985). Taking the easy way out: Non use and misuse of contextual clues in EFL reading comprehension. *English Teaching Forum*, 23(2), 7-20.
- Laufer, B. (1989). What percentage of text lexis is essential for comprehension. In C. Lauren & M. N. Clevedon (Eds.), *Special language: From humans thinking to thinking machines* (pp. 316-323). California: Multilingual Matters.
- Meara, P., & Miralpeix, I. (2016). *Tools for researching vocabulary*. California: Multilingual Matters.
- Moinzadeh, A., & Moslehpour, R. (2012). Depth and breadth of vocabulary knowledge: Which really matters in reading comprehension of Iranian EFL learners? *Journal of Language Teaching and Research*, 3, 1015-1026. doi: 10.4304/jltr.3.5.1015-1026
- Mosback, G., & Mosback, V. (1976). *Practical faster reading: An intermediate/advanced course in reading and vocabulary*. Cambridge: Cambridge University Press.
- Nagy, W. E., Anderson, R. C., & Herman, P. A. (1987). Learning word meanings from context during normal reading. *American Educational Research Journal*, 24(2), 237-270. doi: 10.3102/00028312024002237
- Nation, I. S. P., & Hunston, S. (2013). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Paribakht, T. S., & Wesche, M. (1993). Reading comprehension and second language development in a comprehension-based ESL program. *TESL*

*Canada Journal*, 11(1), 9-29. doi:  
10.18806/tesl.v11i1.623

- Qian, D. (1999). Assessing the role of depth and breadth of vocabulary knowledge in reading comprehension. *The Canadian Modern Language Review*, 56, 282-308. doi: 10.1111/1467-9922.00193
- Rashidi, N., & Khosravi, N. (2010). Assessing the role of depth and breadth of vocabulary knowledge in reading comprehension of Iranian EFL learners. *Journal of Pan-Pacific Association of Applied Linguistics*, 14(1), 81-108.
- Staehr, L. S. (2008). Vocabulary size and the skills of listening, reading and writing. *Language Learning Journal*, 36(2), 139–152. doi: 10.1080/09571730802389975
- Stahl, S. A., Jacobson, M. G., Davis, C. E., & Davis, R. L. (1989). Prior knowledge and difficult vocabulary in the comprehension of unfamiliar text. *Reading Research Quarterly*, 24(1), 27-43. doi: 10.2307/748009
- Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *Modern Language Journal*, 95, 26–43. doi: 10.1111/j.1540-4781.2011.01146.x
- Teng, F. (2018). Incidental vocabulary acquisition from reading-only and reading-while-listening: A multi-dimensional approach. *Innovation in Language Learning and Teaching*, 12(3), 274-288. doi: 10.1080/17501229.2016.1203328