A STUDY ON SECOND- SEMESTER STUDENTS' VOCABULARY SIZE AT ENGLISH STUDY PROGRAM

Dewi Alfiana[1], Indah Tri Purwanti[2], Mahdum[3]

[1, 2, 3] **Universitas Riau, Pekanbaru, Indonesia**

Email: [1]alfiana.dewi1091@student.unri.ac.id, [2]indah.tri@lecturer.unri.ac.id, [3]mahdum@lecturer.unri.ac.id

ABSTRACT

This study examines the receptive vocabulary size of 74 second-semester students in the English Study Program at Universitas Riau using the 140-item Vocabulary Size Test (Nation & Beglar, 2007). Administered online, the test yielded a high reliability (KR-20 = 0.94). Scores ranged from 24 to 129 correct responses (≈2,400–12,900 word families), with a mean of 73 items (≈7,300 word families; SD = 30.93), placing the average student in the "Large" category. Distribution showed 20.3% in "Very Large" (10,001–14,000), 35.1% in "Large" (7,001–10,000), 12.2% in "Medium" (5,001–7,000), and 32.4% in "Small" (2,001–5,000). While most students possess adequate vocabulary for general and academic comprehension, nearly one-third fall below minimal thresholds for 95% text coverage. The study recommends differentiated instruction through regular vocabulary quizzes, extensive academic reading, and peer-collaborative activities. Future research should track learners longitudinally and explore influences on vocabulary growth, such as reading habits, motivation, and teaching strategies.

Keywords: receptive vocabulary size, second semester student', english study program

STUDI TENTANG KEMAMPUAN KOSAKATA MAHASISWA SEMESTER KEDUA PADA PROGRAM STUDI BAHASA INGGRIS

ABSTRAK

Penelitian ini menguji ukuran kosakata reseptif dari 74 mahasiswa semester kedua Program Studi Bahasa Inggris di Universitas Riau menggunakan Tes Ukuran Kosakata (Vocabular Size Test) yang terdiri dari 140 butir soal (Nation & Beglar, 2007). Dilaksanakan secara daring, tes ini menghasilkan reliabilitas yang tinggi (KR 20 = 0,94). Skor berkisar antara 24 hingga 129 respons yang benar (≈2.400–12.900 keluarga kata), dengan rata-rata 73 butir soal (≈7.300 keluarga kata; SD = 30,93), yang menempatkan rata-rata mahasiswa dalam kategori "Besar". Distribusi menunjukkan 20,3% dalam "Sangat Besar" (10.001–14.000), 35,1% dalam "Besar" (7.001–10.000), 12,2% dalam "Sedang" (5.001–7.000), dan 32,4% dalam "Kecil" (2.001–5.000). Sementara sebagian besar siswa memiliki kosakata yang memadai untuk pemahaman umum dan akademis, hampir sepertiga berada di bawah ambang batas minimal untuk cakupan teks 95%. Studi ini merekomendasikan instruksi yang dibedakan melalui kuis kosakata reguler, bacaan akademis yang ekstensif, dan aktivitas kolaboratif sebaya. Penelitian di masa mendatang harus melacak pelajar secara longitudinal dan mengeksplorasi pengaruh pada pertumbuhan kosakata, seperti kebiasaan membaca, motivasi, dan strategi pengajaran.

Kata Kunci: ukuran kosakata reseptif, mahasiswa semester dua, program studi bahasa inggris

INTRODUCTION

The growing role of English as an international lingua franca has led Indonesian curricula from primary through tertiary levels to emphasize English instruction (Dardjowidjojo, 2002; Sikki et al., 2013; Ivone, 2005). Yet despite six years of formal study, many Indonesian learners continue to struggle with low proficiency, often attributable to limited vocabulary knowledge. Schmitt (2000) further underscores that vocabulary is essential for listening, speaking, reading and writing without which communication and academic success are hindered.

Research indicates that adequate text comprehension requires holding between 3,000 and 5,000 word families for 95 percent coverage of authentic texts, with optimal thresholds of 8,000 and 9,000 for 98 percent coverage (Nation, 2006; Laufer and Ravenhorst Kalovski, 2010). For spoken discourse, 2,000 to 3,000 high frequency word families suffice, whereas 6,000 and 7,000 word families are needed to navigate academic listening (Adolphs and Schmitt, 2003; Nation, 2006). These benchmarks guide curriculum designers in setting realistic vocabulary targets for EFL learners.

Several Indonesian studies have measured vocabulary size among undergraduates. Quinn (1968) found entrants knew approximately 2,800 word families; Nurweni and Read (1999) and Nurhemida (2007) reported similarly modest figures; and Kweldju (1997) observed averages below 5,000 word families. However, most research has focused on senior or mixed level cohorts, leaving lower level students particularly second semester English majors underrepresented in the literature.

To fill this gap, the present study measures the receptive vocabulary size of second semester students in the English Study Program at Universitas Riau using the validated 140 item Vocabulary Size Test (Nation and Beglar, 2007). By mapping individual scores onto established frequency bands and CEFR aligned thresholds, this research aims to provide baseline data for differentiated vocabulary instruction and to inform both lecturers and curriculum developers on strategies to enhance lexical growth in the early phase of tertiary study.

LITERATURE REVIEW

In the field of second language acquisition, vocabulary knowledge is regarded as the cornerstone of communicative competence, underpinning listening, speaking, reading, and writing (Milton & Daller, 2013; Wilkins, 1972 as cited in AlfAki, 2014) . Without sufficient lexical breadth, learners—even with strong grammatical knowledge—struggle to convey or comprehend meaning (Schmitt & Schmitt, 2014) . Read (1997 as cited in Hajiyeva, 2015) further emphasizes that accurate assessment of vocabulary size is essential for tailoring instruction and tracking learner progress, since vocabulary breadth both enables and is reinforced by language use.

Researchers classify vocabulary into high-frequency, academic, technical, and low-frequency words, each playing distinct roles in language use (Nation, 2008) . High-frequency words (the most common 2,000 word families) account for roughly 80 percent of running words in written texts and 90 percent in spoken discourse, serving as the foundation for everyday communication. Academic words (as in Coxhead's Academic Word List) and technical terms support comprehension of specialized texts, while low-frequency vocabulary enables nuanced expression. Empirical benchmarks suggest that 2,000–3,000 word families suffice for basic oral communication (Adolphs & Schmitt, 2003), 6,000–7,000 for academic listening (Nation, 2006), and 3,000–5,000 for 95 percent coverage of unsimplified texts, with 8,000–9,000 needed for 98 percent coverage (Laufer & Ravenhorst-Kalovski, 2010).

In the Indonesian context, foundational studies reveal that university entrants often possess limited vocabularies. Quinn (1968) reported an average of 2,800 word families among new undergraduates; Nurweni and Read (1999) observed that English education majors knew only about 2,800 word families . Subsequent investigations by Kurniawan (2017) and others confirm that many undergraduates fall below the thresholds recommended for effective academic comprehension, highlighting a persistent gap between curriculum aims and learner lexicons.

More recent empirical work continues to document these gaps. Wero, et al, (2021) found that eighth-semester English students averaged just over 3,100 receptive word families (and 1,841 productive. Siregar (2019) reported an average of 8,732 word families among second-semester students at a private West Java university, yet noted that 75 percent mastered only a limited

high- and mid-frequency vocabulary band . Sudarman and Sumalee (2018) observed mean receptive sizes of - 1,273 word families among first-year students at Kutai Kartanegara University, and Kirana and Basthomi (2020) found averages under 1,400 word families at IAIN Ponorogo . Risky et al, (2022) added that factors such as instructional methods and learner motivation further influence vocabulary growth . Together, these theoretical and empirical insights underscore the necessity of measuring receptive vocabulary size among second-semester undergraduates to inform targeted, differentiated instruction in Indonesian EFL settings.

.

RESEARCH METHOD

This study was carried out at the English Study Program, Teacher Training and Education Faculty, Universitas Riau, located on Bina Widya Campus, KM 12.5 Simpang Baru, Pekanbaru, Indonesia, during the 2024–2025 even semester. Adopting a descriptive-quantitative design, the research aimed to obtain a clear, numerical picture of second-semester students' receptive vocabulary size without manipulating variables or testing causal hypotheses. Quantitative descriptive research analyzes and presents data in numerical form, focusing on actual phenomena to produce objective and verifiable results .

The research procedure unfolded in five key stages: preparation of the proposal and review of relevant literature; purposive sampling of participants; adaptation and validation of the Vocabulary Size Test (VST); online administration of the instrument; and scoring and statistical analysis of responses.The population comprised all second-semester students in classes B and C of the English Study Program, academic year 2024/2025, while class A (39 students) served as the try-out sample for instrument validation . A total of 74 students participated in the main study, meeting criteria of active enrollment and voluntary consent. The instrument was the full 140-item Vocabulary Size Test developed by Nation and Beglar (2007), covering 14 frequency levels from the 1,000th to the 14,000th most frequent word families. Each correct response approximates knowledge of 100 word families.

Data were collected online via Google Forms distributed through WhatsApp groups. No strict time limit or technical instructions were imposed; participants completed the test independently and honestly, and responses were automatically compiled for analysis. The data analysis process is carried out in MS Excel. Analysis employed descriptive statistics: calculation of mean, highest and lowest scores, and standard deviation; conversion of raw scores into estimated vocabulary size (word families); and categorization into proficiency bands (Very Small to Very Large). The vocabulary size categories can be seen in table 1 as follows:

Table 1. Vocabulary Size Categories

Category	Vocabulary Size (Word Families)	Description	Typical Word Frequency Range
Very Small	Less than 2,000	Basic everyday vocabulary	High-frequency words only
Small	2,000 - 5,000	Basic plus some less frequent words	Mostly high- and mid-frequency words
Medium	5,001 - 7,000	Intermediate vocabulary size	Includes high-, mid-, and some low-frequency words
Large	7,001 - 10,000	Advanced vocabulary size	High, mid, and low frequency words
Very Large	10.001 - 14,000	Near-native or native-like mastery	Extensive range including rare words

Notion and Beglar, (2007)

RESULTS AND DISCUSSION

This section aims to answer the research question regarding the extent of students' vocabulary knowledge and how their vocabulary size is categorized based on the results of the Vocabulary Size Test (VST). The data were obtained from students' responses to 140 multiple-choice items adapted from Nation and Beglar's (2007) Vocabulary Size Test, which covers 14 frequency levels ranging from the 1,000-word family level to the 14,000-word family level. The total number of correct answers was then converted into an estimated vocabulary size (in word families), and each student's result was categorized into proficiency levels such as beginner, intermediate, advanced, or near-native based on a vocabulary size classification scheme.

1. Validity and Reliability Test

According to the results of the validity test using a significance value (α) = 0.05 along with a total of 39 correspondents, the rtable value was 0.325. From the results of the analysis carried out using MS Excel, there were 99 valid questions out of a total of 140 questions. This states that 70.7% of the questions are valid. For the reliability test value obtained was 0.94. Wombacher, 2017) states that the instrument is reliable if reliability coefficient is more than 0.80 with Kuder-Richardson Formula 20 (KR-20). From this statement, the instrument can be used in research. The results of the reliability test can be seen in table 2.

Table 2. Reliability Test Result

Realibility Test	Σ Responden	Σ Question	Varians total	Reliability Coefficient	Reliability Categories
KR - 20	39	140	377,887	0,942	Very High

2. Vocabulary Test Result

Table 3. Vocabulary Size Test Result

No	Vocab Size	Nuber of Student	Categories	Percentage
1	10.001 - 14.000	15	Very Large	20,27%
2	7.001- 10.000	26	Large	35,14%
3	5,001 - 7.000	9	Medium	12,16%
4	2.001 - 5.000	24	Small	32,43%

The analysis shows that the majority of participants (26 individuals) fall into the *Large* category, suggesting a vocabulary size between 7,000 to 9,000 word families. According to Beglar (2010), this vocabulary range is generally sufficient to understand about 95% of the vocabulary in common texts, enabling learners to engage in general communication and comprehend simpler academic materials, though more advanced comprehension might still be challenging. Meanwhile, 24 participants are in the *Small* category, indicating mastery of fewer than 3,000 word families. This group likely faces difficulties understanding academic texts or authentic English materials. Research by Stoeckel, Ishikawa, and Bennett (2013) suggests that learners in this category often cannot comprehend more than 80–85% of words in a text, which significantly limits overall understanding and academic engagement.

Participants reached the *Very Large* category (approximately 10,000–14,000 word families), representing near-native vocabulary knowledge. Learners at this level are expected to be capable of understanding rare and specialized vocabulary found in advanced academic texts and professional discourse. This level reflects extensive exposure to English through reading and listening.

Finally, only 9 participants were classified in the *Medium* category, mastering around 3,000–6,000 word families. This indicates moderate proficiency, where learners can understand general content but may struggle with more complex or specialized texts.

3. Mean, Standart Deviation, Highest Score and Lowest Score

Table 4. Mean, Standart Deviation, Highest Score, Lowest Score

Mean	Standart Deviation	Highest Score	Lowest Score
73	30,930	129	24

The descriptive statistics from the Vocabulary Size Test (VST) indicate that the mean score of participants was 73 out of 140 items, suggesting a moderate level of vocabulary knowledge across the group. The standard deviation was relatively high at 30.93, implying a considerable spread in the data and substantial variation in students' vocabulary sizes Gravetter & Wallnau (2014). The highest score achieved was 129, while the lowest was 24, demonstrating a wide range of proficiency levels among participants. The formulation in the form of a table related to the standard deviation value, average and highest/lowest score can be seen in the following table.

This study found that second-semester students achieved a mean score of 73 out of 140 ($\approx$7,300 word families), placing them in the "Large" category and indicating sufficient receptive vocabulary for general and academic texts. This result aligns closely with Mahsyur (2019), who reported averages below 10,000 word families, and is slightly lower than Siregar's (2019) finding of 8,732.5 word families, despite differences in instruments and institutional contexts. In contrast, Wero et al. (2021) observed much smaller vocabularies (3,110 receptive; 1,841 productive) among eighth-semester students, likely due to their use of RVST/PVST tests and cohort differences in English exposure.

The lowest score of 24 ($\approx$2,400 word families) still meets the threshold for basic oral communication (Adolphs & Schmitt, 2003), while the highest score of 129 ($\approx$12,900) approaches near-native levels, encompassing low-frequency and specialized vocabulary (Nation & Beglar, 2007). Most participants (26) fell within the 7,000–9,000 band sufficient for about 95 percent text coverage (Beglar, 2010), but 24 students remained below 3,000 word families and may struggle with authentic materials.

Mapped against CEFR benchmarks, the cohort's mean of 7,300 word families corresponds to B2–C1 proficiency, enabling advanced academic engagement. Nevertheless, to achieve optimal comprehension 95 % coverage at 4,000–5,000 word families and 98% at 8,000–9,000. further vocabulary development is recommended (Nation, 2006; Laufer & Ravenhorst-Kalovski, 2010).

RESULTS AND DISCUSSION

The study revealed that second-semester students in the English Study Program at Universitas Riau possess a mean receptive vocabulary size of approximately 7,300 word families, classifying them within the "Large" category; individual scores, however, ranged from 2,400 to 12,900 word families, indicating substantial variability in lexical knowledge among participants. While around 41 percent of students achieved vocabulary sizes sufficient for 95–98 percent coverage of academic texts, roughly 36 percent remained below minimal thresholds, underscoring the need for differentiated instructional support.

It is therefore recommended that lecturers integrate vocabulary development into discipline-relevant reading and tasks, administer regular quizzes with personalized feedback, and monitor individual growth trajectories. Students should commit to daily academic reading, record and contextualize new lexis, and participate in peer study groups to reinforce retention.

Future investigations might employ advanced statistical modeling and richer field-based data such as digital reading logs or classroom discourse recordings to elucidate factors influencing vocabulary acquisition and to inform more targeted pedagogical interventions.

REFERENCES

Adolphs, S., & Schmitt, N. (2003). Lexical coverage of spoken discourse. *Applied Linguistics, 24*(4), 425–438.

Alfaki, I. M. (2014). Vocabulary input in EFL textbooks: A comparative study of vocabulary in two EFL textbooks. *English Language Teaching, 7*(3), 87–96.

Beglar, D. (2010). A Rasch-based validation of the Vocabulary Size Test. *Language Testing, 27*(1), 101–118.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*(2), 213–238.

Dardjowidjojo, S. (2002). English teaching in Indonesia. *EA Journal, 20*(1), 22–30.

Gravetter, F. J., & Wallnau, L. B. (2014). *Essentials of statistics for the behavioral sciences* (8th ed.). Wadsworth, Cengage Learning.

Hajiyeva, K. (2015). Exploring the relationship between receptive vocabulary knowledge and academic achievement of EFL learners. *English Language Teaching, 8*(8), 1–10.

Ivone, F. M. (2005). Extensive listening for EFL students: A study of listening fluency development. *TEFLIN Journal, 16*(2), 85–96.

Kirana, M. B., & Basthomi, Y. (2020). Vocabulary knowledge of Indonesian EFL university students: A study on receptive vocabulary size. *Language Circle: Journal of Language and Literature, 14*(2), 205–214.

Kurniawan, D. (2017). Measuring the vocabulary size of Indonesian EFL university students using Nation's vocabulary size test. *Jurnal Ilmu Pendidikan, 23*(1), 27–34.

Kweldju, S. (1997). English vocabulary size and the vocabulary learning strategies of Indonesian university students. *TEFLIN Journal, 8*(1), 1–18.

Laufer, B., & Ravenhorst-Kalovski, G. C. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language, 22*(1), 15–30.

Mahsyur, M. (2019). An analysis of students' vocabulary size at IAIN Palopo. *Jurnal Edulingua, 6*(2), 1–12.

Milton, J., & Daller, H. (2013). *Vocabulary size revisited: The link between vocabulary size and academic achievement*. Multilingual Matters.

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review, 63*(1), 59–82.

Nation, I. S. P. (2008). *Teaching vocabulary: Strategies and techniques*. Heinle Cengage Learning.

Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher, 31*(7), 9–13.

Nurhemida. (2007). Measuring the vocabulary size of English education students at Riau University. *Jurnal Ilmu Pendidikan, 12*(1), 20–30.

Nurweni, A., & Read, J. (1999). The English vocabulary knowledge of Indonesian university students. *English for Specific Purposes, 18*(2), 161–175.

Quinn, A. (1968). English vocabulary levels of Indonesian university entrants. *RELC Journal, 1*(1), 48–55.

Risky, R. T., Haryanto, E., & Sutrisno, A. (2022). The influence of motivation and learning strategy on vocabulary acquisition. *Jurnal Bahasa dan Sastra, 11*(2), 145–157.

Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge University Press.

Schmitt, N., & Schmitt, D. (2014). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching, 47*(4), 484–503.

Sikki, A., Rahman, M. A., Hamra, A., & Noni, N. (2013). Developing a model of teaching reading comprehension based on problem solving. *International Journal of Linguistics, 5*(1), 1–12.

Siregar, R. M. (2019). An analysis of vocabulary size of university students in West Java. *ELTIN Journal, 7*(1), 30–42.

Stoeckel, T., Ishikawa, S., & Bennett, P. (2013). Vocabulary size and the vocabulary levels test. *System, 41*(3), 663–672.

Sudarman, A., & Sumalee, S. (2018). Vocabulary knowledge of first-year students at Kutai Kartanegara University. *Jurnal Pendidikan dan Kebudayaan, 23*(3), 187–194.

Wero, D. F., Hendrawaty, N., & Aksa, A. (2021). Receptive and productive vocabulary size of Indonesian EFL learners. *Jurnal Pendidikan Bahasa Inggris, 9*(1), 15–23.

Wilkins, D. A. (1972). *Linguistics in language teaching*. Edward Arnold.

Wombacher, J. (2017). Reliability assessment in psychology: An introduction. *Journal of Psychology and Psychotherapy, 7*(4), 1–5.

Dewi Alfiana, Indah Tri Purwanti, Mahdum | receptive vocabulary size, second semester student', english study program

Page | 199