

Vocabulary Mastery and TOEFL ITP Listening Performance: Correlational Evidence from Indonesian EFL Undergraduates

Received: April 11, 2026; Revised: May 11, 2026; Accepted: 11 June 2026

Nyoman Yuniantari^{1*}, Sri Rahayu²

¹²Cokroaminoto Palopo University, Indonesia

✉ * Corresponding author: kerti2406@gmail.com

Abstract: Vocabulary knowledge is considered an important linguistic resource for understanding spoken English, yet its relationship with TOEFL ITP listening performance remains underexamined among Indonesian undergraduates approaching program completion. This study investigated the relationship between vocabulary mastery and TOEFL ITP listening performance among seventh-semester students in the English Education Study Program at Cokroaminoto Palopo University. A quantitative correlational design was employed, involving 17 students selected purposively from the reported population of 108 students. Data were collected using a 30-item multiple-choice vocabulary mastery test and a 50-item listening test covering short conversations, extended conversations, and talks. Listening scores were converted to a 1–100 scale before descriptive classification. The data were analyzed using descriptive statistics, the Shapiro–Wilk normality test, an analysis-of-variance test for linearity, and a two-tailed Pearson product-moment correlation. Vocabulary mastery was predominantly classified as fair, with a mean of 63.18 (SD = 12.39), whereas listening performance was predominantly poor or very poor, with a mean of 38.82 (SD = 15.26). The variables were positively and significantly correlated, $r(15) = .669$, $p = .003$, 95% CI (.277, .870), and shared approximately 44.7% of their variance. These findings provide context-specific evidence that stronger vocabulary mastery is associated with better TOEFL ITP listening performance, while not establishing causality. Vocabulary instruction should therefore incorporate auditory word recognition and contextualized listening practice. Results also underline the distinction between lexical knowledge and real-time auditory access. Because the sample was small and institution-specific, larger multisite studies using validated aural vocabulary measures and authorized standardized listening instruments are recommended.

Keywords: correlational evidence, EFL undergraduates, listening performance, TOEFL, vocabulary mastery.

1 Introduction

Listening comprehension is central to successful participation in English-medium higher education because students must identify propositions, infer speakers' intentions, and integrate information while spoken input unfolds only once. Research on academic vocabulary likewise shows that university learners require sufficiently broad and discipline-sensitive lexical resources to engage with lectures, assessments, and other academic discourse, yet many perceive academic words as complex, infrequent outside university, and difficult to retain (Chung et al., 2024; Lu & Dang, 2022). Standardized listening tests intensify these demands by

combining real-time processing with restricted response time, shifting topics, and vocabulary distributed across conversations and extended talks; corpus analyses indicate that substantial lexical coverage is necessary for adequate comprehension of test passages and academic spoken English, although required coverage varies with genre, task, and learner characteristics (Phung & Ha, 2022; Trang et al., 2023). Consequently, vocabulary knowledge deserves attention not merely as an instructional outcome but as a plausible linguistic foundation for performance in academically oriented listening assessments.

Vocabulary is multidimensional: learners may recognize a written form without retrieving its spoken form rapidly, understand an isolated word without interpreting it in connected discourse, or know a general meaning without accessing contextually appropriate senses. Empirical work with university EFL learners confirms that phonological vocabulary knowledge is strongly associated with listening comprehension, whereas aural and orthographic vocabulary measures represent related but non-equivalent capacities and may contribute differently across proficiency levels and listening tasks (Masrai, 2022; Hamada & Yanagawa, 2024). Listening success also depends on converting a continuous acoustic signal into recognizable lexical units while allocating limited attention to syntax, propositions, and discourse meaning; failures in word segmentation, connected-speech perception, or rapid lexical access can therefore disrupt comprehension even when words appear familiar in print (Ke & Wang, 2022; Wong et al., 2021). Investigating vocabulary alongside listening achievement is thus theoretically warranted, particularly where instruction and testing may privilege written knowledge while assessment requires immediate recognition of spoken language.

Within Indonesian tertiary EFL settings, TOEFL ITP scores commonly function as indicators of English readiness and may influence students' academic progression, yet listening remains difficult when learners must follow short exchanges, longer conversations, and monologic talks under single-play conditions. Broader assessment research demonstrates that performance across receptive modalities is not interchangeable and that the linguistic knowledge supporting one modality may manifest differently in another, making institution-specific evidence necessary before instructional priorities are established (He et al., 2022; Fung & Chung, 2024). The central concern is whether students who possess stronger vocabulary mastery also attain higher TOEFL listening scores, rather than whether vocabulary alone causes listening success, because listening performance simultaneously reflects linguistic knowledge, metacognitive regulation, background knowledge, attention, and task processing. Models incorporating such person and strategy factors nevertheless identify aural vocabulary size as a significant predictor, while reviews of processing problems consistently position lexical recognition among the principal constraints on L2 comprehension (Du & Man, 2022; Goh, 2023).

A defensible response to this concern is to examine the association between measured vocabulary mastery and standardized listening achievement within the population for whom the assessment carries direct academic relevance. Correlational evidence cannot establish causality, but it can determine the direction and magnitude of covariation, identify whether vocabulary merits inclusion in subsequent explanatory models, and provide an empirical basis for locally targeted instruction; recent studies using auditory vocabulary measures have repeatedly found meaningful positive relationships with L2 listening performance (Du et al., 2022; Wang, 2023a). Such analysis should preserve the distinction between lexical knowledge and listening achievement while recognizing that both are measured constructs whose scores depend on item content, response format, and modality. Research comparing aural vocabulary breadth and depth indicates that breadth may explain more listening variance, whereas work on productive vocabulary demonstrates that accessible lexical representations can accelerate sentence-level comprehension, supporting a relational hypothesis rather than a causal claim (Han & Qian, 2024; Allal-Sumoto et al., 2023).

The proposed relationship can be addressed pedagogically by strengthening learners' ability to connect word forms, meanings, and uses with the phonological realizations

encountered in continuous spoken speech. Because written familiarity does not guarantee immediate auditory recognition, instruction should integrate repeated listening, form-focused noticing, contextualized meaning retrieval, and practice with connected speech; studies of aural vocabulary levels and decoding suggest that rapid access to high-frequency spoken words provides essential support for constructing message-level meaning (Wang, 2023b; Giordano, 2023). Exposure should also reflect academic listening conditions rather than rely exclusively on isolated word study, enabling learners to encounter vocabulary within lectures, conversations, and information-dense monologues. Investigations of academic lectures show that spoken input can generate incidental lexical gains and that words and collocations are learned differently across listening, reading, viewing, and combined modes, thereby supporting deliberate integration of vocabulary development with authentic comprehension activity (Reynolds et al., 2022; Dang et al., 2022a).

Multimodal and strategy-supported practice offers a complementary route because transcripts, captions, and controlled replay can temporarily reduce perceptual burden while learners establish stronger, readily accessible sound–meaning mappings during sustained, authentic, and progressively demanding academic listening practice. Captioned viewing has produced gains in several dimensions of word knowledge, and comparisons across input modes show that captions and reading-while-listening may support recognition and retention more effectively than listening alone, although outcomes remain moderated by proficiency, prior vocabulary knowledge, and working memory (Teng, 2022; Teng, 2024). Technology should nevertheless scaffold rather than replace unaided listening, since standardized tests ultimately require learners to process speech without textual support and within fixed time constraints. Experimental evidence indicates that caption availability and playback speed interact with proficiency and item difficulty, while structured podcast work can improve both listening comprehension and vocabulary, especially when learners collaborate around meaning-focused tasks (Mahalingappa et al., 2024; Saeedakhtar et al., 2021).

Closely related studies consistently connect vocabulary with listening, but their operationalizations and contexts leave a consequential gap. Research has associated receptive and productive phonological vocabulary with listening performance, linked aural and orthographic phrasal-verb knowledge to comprehension, and shown that learners' linguistic knowledge interacts with their strategy use; however, these studies predominantly involve Arabic-, Chinese-, or Japanese-speaking learners and often employ general listening or TOEIC measures rather than TOEFL ITP achievement among Indonesian English majors (Kang, 2024; Wakamoto & Rose, 2021). Further intervention research demonstrates that viewing an academic lecture can facilitate incidental learning of single words and collocations, while validation studies show that corpus-derived word lists and the definition of a "word" influence vocabulary-test interpretations; these advances clarify possible mechanisms but do not establish whether a conventional vocabulary mastery score is associated with locally administered TOEFL ITP listening performance in a small, academically consequential cohort (Dang et al., 2022b; Pinchbeck et al., 2022).

Accordingly, this study aims to determine whether vocabulary mastery is significantly correlated with TOEFL ITP listening achievement among seventh-semester students in the English Education Study Program at Cokroaminoto Palopo University. Its novelty lies in bringing a vocabulary–listening relationship that has largely been examined through general listening, TOEIC, IELTS, or experimentally designed measures into a locally consequential TOEFL ITP context involving students approaching program completion. The hypothesis is that vocabulary mastery and TOEFL ITP listening achievement are positively associated because learners with more accessible lexical knowledge should recognize a larger proportion of spoken items and devote fewer processing resources to word identification. The inquiry is deliberately restricted to the direction, strength, and statistical significance of the relationship between the two measured variables in the participating cohort; it does not test instructional effects, claim

that vocabulary causes listening achievement, or estimate the independent contributions of pronunciation, working memory, metacognitive knowledge, anxiety, accent familiarity, test-taking strategy, and prior TOEFL experience.

2 Methods

This study employed a quantitative correlational design to examine the relationship between vocabulary mastery and TOEFL ITP listening performance. The research was conducted in the language laboratory of Cokroaminoto Palopo University, Indonesia, on April 6–7, 2026. The target population comprised students enrolled in the English Education Study Program during the 2025/2026 academic year, which was reported to include 108 students across four semester cohorts. Participants were selected through purposive sampling according to three criteria: they were active students in the study program, had previously studied TOEFL materials, particularly the listening section, and voluntarily agreed to participate. Seventeen seventh-semester students satisfied these criteria and constituted the final sample. The selection of students at this stage of their undergraduate program was relevant because they had completed various English-language courses, had been introduced to the TOEFL format, and were approaching the academic stage at which TOEFL performance was required by their institution.

Data were collected using a 30-item multiple-choice vocabulary mastery test and a 50-item TOEFL ITP listening test. The items were adapted from publicly available instructional materials, and their content was reviewed in consultation with an English-language teaching and testing specialist. Before each test, the researcher explained the procedures and distributed answer sheets to the participants. The vocabulary questions were presented sequentially on an LCD screen, with each item displayed for 10 seconds, and the test session was completed within 10 minutes. The listening test was administered through loudspeakers in the language laboratory, while the questions were displayed on an LCD screen. It comprised three parts corresponding to the TOEFL ITP listening format: Part A contained 30 short-conversation items and lasted 15 minutes, Part B contained eight extended-conversation items and lasted seven minutes, and Part C contained 12 talk-based items and lasted 13 minutes. Participants selected one response from four options without external assistance.

Following data collection, the vocabulary and listening responses were scored and entered into IBM SPSS Statistics version 26. TOEFL listening section scores, originally expressed on a scale from 31 to 68, were converted to an academic scale from 1 to 100 using the formula: $\text{academic score} = [(\text{TOEFL listening score} - 31)/37] \times 99 + 1$. Descriptive classifications were subsequently applied to both variables: scores of 76–100 were categorized as good, 56–75 as fair, 40–55 as poor, and 1–39 as very poor. Before hypothesis testing, the distributions of the two variables were examined using the Shapiro–Wilk test because the sample comprised fewer than 50 participants. Linearity was evaluated using the significance value for deviation from linearity. After the assumptions of normality and linearity had been assessed, a two-tailed Pearson product-moment correlation was calculated to determine the direction, strength, and statistical significance of the relationship between vocabulary mastery and TOEFL ITP listening performance, with statistical significance established at $\alpha = .05$.

3 Results

This section presents the findings concerning vocabulary mastery and TOEFL ITP listening performance among the participating Indonesian EFL undergraduates. The analysis begins with descriptive statistics and performance classifications for both variables. It is followed by preliminary assumption testing and an examination of the direction, strength, and statistical

significance of the relationship between vocabulary mastery and TOEFL ITP listening performance.

3.1 Descriptive Results for Vocabulary Mastery and TOEFL ITP Listening Performance

Descriptive analysis was conducted to summarize the central tendency, dispersion, observed range, and performance classifications of vocabulary mastery and TOEFL ITP listening performance among the 17 participants. Vocabulary mastery was measured using a 30-item multiple-choice test, whereas listening performance was measured using a 50-item test representing Parts A, B, and C of the TOEFL ITP listening section. For comparability in descriptive reporting, the listening scores were converted from the original TOEFL section scale of 31–68 to an academic scale of 1–100 before classification. Scores from 76 to 100 were categorized as good, scores from 56 to 75 as fair, scores from 40 to 55 as poor, and scores below 40 as very poor. The descriptive analysis included the mean, standard deviation, minimum score, maximum score, frequency, and percentage for each performance category. Table 1 summarizes these statistics and shows the distribution of participants across the four predefined levels for both variables.

Table 1: Descriptive Statistics and Performance Classifications for Vocabulary Mastery and TOEFL ITP Listening Performance

Variable	<i>n</i>	<i>M</i>	<i>SD</i>	Minimum	Maximum	Good, <i>n</i> (%)	Fair, <i>n</i> (%)	Poor, <i>n</i> (%)	Very poor, <i>n</i> (%)
Vocabulary mastery	17	63.18	12.39	40	80	4 (23.5)	9 (52.9)	4 (23.5)	0 (0.0)
TOEFL ITP listening performance	17	38.82	15.26	12	71	0 (0.0)	2 (11.8)	5 (29.4)	10 (58.8)

Note. *M* = mean; *SD* = standard deviation. TOEFL ITP listening scores were converted to a 1–100 scale before descriptive classification.

As shown in Table 1, vocabulary mastery scores ranged from 40 to 80, producing an observed range of 40 points. The participants obtained a mean score of 63.18 with a standard deviation of 12.39, placing the average vocabulary result within the fair performance category. Nine of the 17 participants, representing 52.9% of the sample, achieved scores between 56 and 75 and were therefore classified as fair. Four participants, or 23.5%, obtained scores from 76 to 100 and were classified as good. The remaining four participants, also representing 23.5% of the sample, recorded scores between 40 and 55 and were categorized as poor. No participant obtained a vocabulary score below 40; consequently, none was placed in the very poor category. Overall, the distribution was concentrated in the middle classification, although the difference between the minimum and maximum scores indicates substantial and noticeable variation in vocabulary mastery among the participating seventh-semester students.

The converted TOEFL ITP listening scores showed a different distribution, ranging from 12 to 71 and yielding an observed range of 59 points. The mean listening score was 38.82, with a standard deviation of 15.26, which placed the average performance within the very poor category. Ten participants, accounting for 58.8% of the sample, scored below 40 and were categorized as very poor. Five participants, representing 29.4%, obtained scores between 40 and 55 and were classified as poor, whereas only two participants, or 11.8%, achieved scores between 56 and 75 and reached the fair category. None of the participants recorded a listening

score of 76 or higher; therefore, the good category contained no cases. The larger standard deviation and observed range for listening also demonstrate greater variability than that found for vocabulary scores. Collectively, the distribution indicates that most participants' listening results were concentrated in the two lowest categories, despite their predominantly fair vocabulary mastery.

3.2 Assessment of Normality

Before estimating the Pearson product-moment correlation, the distributional form of each variable was examined to determine whether the normality assumption was reasonably satisfied. The Shapiro-Wilk test was selected because the analysis involved a small sample of 17 participants and provided a formal assessment of whether the observed score distributions differed significantly from a normal distribution. The test was applied separately to the vocabulary mastery scores and the converted TOEFL ITP listening scores. Statistical significance was evaluated at an alpha level of .05. Under this decision rule, a probability value below .05 would indicate statistically significant evidence that a distribution departed from normality, whereas a probability value above .05 would provide insufficient evidence to reject the assumption of normality. The analysis focused on the Shapiro-Wilk statistic, degrees of freedom, and corresponding probability value for each variable. The complete numerical test results of this preliminary assumption test are reported in Table 2.

Table 2: Shapiro-Wilk Normality Test for Vocabulary Mastery and TOEFL ITP Listening Performance

Variable	Shapiro-Wilk statistic	<i>df</i>	<i>p</i>
Vocabulary mastery	.931	17	.230
TOEFL ITP listening performance	.963	17	.679

Note. TOEFL ITP listening performance refers to the scores converted to a 1–100 scale. The significance criterion was $\alpha = .05$.

As presented in Table 2, the vocabulary mastery scores produced a Shapiro-Wilk statistic of .931 with 17 degrees of freedom and a probability value of .230. Because this probability value exceeded the predefined alpha level of .05, the null hypothesis of a normally distributed vocabulary score pattern was not rejected. The converted TOEFL ITP listening scores yielded a Shapiro-Wilk statistic of .963 with 17 degrees of freedom and a probability value of .679. This value also exceeded .05, indicating insufficient evidence to reject normality for the listening scores. Accordingly, neither variable demonstrated a statistically significant departure from a normal distribution according to the Shapiro-Wilk test. These findings supported continued evaluation of the assumptions required for Pearson correlation. However, the nonsignificant results should not be interpreted as definitive proof of normality, particularly because the sample was small; they indicate only that the test detected no significant distributional violation in the available observations.

3.3 Assessment of Linearity

Before calculating Pearson's correlation coefficient, the linearity of the relationship between vocabulary mastery and converted TOEFL ITP listening performance was assessed. This assessment was necessary because Pearson's product-moment correlation summarizes the strength of a linear association and may provide an incomplete representation when the relationship between two variables is substantially nonlinear. An analysis-of-variance test for

linearity was therefore conducted using vocabulary mastery as the grouping variable and listening performance as the outcome variable. The relevant decision criterion was the probability value associated with deviation from linearity. A value greater than .05 indicated that the observed departures from a straight-line relationship were not statistically significant, whereas a value below .05 would suggest that the linear specification was inadequate. The analysis also reported the combined between-groups component and the specific linear component of the association. Table 3 presents the sums of squares, degrees of freedom, mean squares, *F* statistics, and probability values obtained from the linearity assessment.

Table 3: Analysis of Variance for the Linearity Assessment

Source of variation	Sum of squares	<i>df</i>	Mean square	<i>F</i>	<i>p</i>
Between groups (combined)	2,091.304	10	209.130	0.768	.661
Linear component	1,665.245	1	1,665.245	6.118	.048
Deviation from linearity	426.059	9	47.340	0.174	.990
Within groups	1,633.167	6	272.194	—	—
Total	3,724.471	16	—	—	—

Note. TOEFL ITP listening performance refers to scores converted to a 1–100 scale. The significance criterion was $\alpha = .05$.

As displayed in Table 3, the linear component was statistically significant, $F(1, 6) = 6.12$, $p = .048$, indicating that vocabulary mastery and converted TOEFL ITP listening performance shared a detectable linear association. More importantly for evaluating the assumption, the deviation-from-linearity component was not statistically significant, $F(9, 6) = 0.17$, $p = .990$. Because the probability value for deviation from linearity was substantially greater than the .05 criterion, the null hypothesis that deviations from a linear pattern were attributable to sampling variation was not rejected. The combined between-groups component was also nonsignificant, $F(10, 6) = 0.77$, $p = .661$; however, this omnibus result did not alter the specific assessment of linearity. Taken together, the significant linear component and nonsignificant deviation component provided no evidence of meaningful nonlinearity in the observed data. The relationship was therefore treated as sufficiently linear to proceed with Pearson's product-moment correlation, while acknowledging the limited precision associated with the small sample.

3.4 Correlation Between Vocabulary Mastery and TOEFL ITP Listening Performance

After the assumptions of approximate normality and linearity had been evaluated, a two-tailed Pearson product-moment correlation was calculated to examine the relationship between vocabulary mastery and converted TOEFL ITP listening performance. The analysis used paired scores from all 17 participants, with vocabulary mastery treated as one continuous variable and listening performance as the second continuous variable. Statistical significance was evaluated at an alpha level of .05. The correlation coefficient was interpreted according to both its direction and absolute magnitude, while the probability value was used to determine whether the observed association was statistically distinguishable from zero. To provide information about estimation precision, a 95% confidence interval for the population correlation was calculated using Fisher's *z* transformation. The squared correlation coefficient was also obtained to describe the proportion of variance shared statistically by the two variables,

without implying that either variable caused the other. Table 4 presents the correlation coefficient, confidence interval, significance level, shared variance, and sample size.

Table 4: Pearson Correlation Between Vocabulary Mastery and TOEFL ITP Listening Performance

Variables	<i>n</i>	<i>r</i>	95% CI for <i>r</i>	<i>p</i>	<i>r</i> ²
Vocabulary mastery and TOEFL ITP listening performance	17	.669	.277, .870	.003	.447

Note. CI = confidence interval. The analysis was two-tailed. TOEFL ITP listening performance refers to scores converted to a 1–100 scale.

As reported in Table 4, vocabulary mastery was positively and significantly correlated with converted TOEFL ITP listening performance, $r(15) = .669$, $p = .003$. The positive coefficient indicates that participants with higher vocabulary scores obtained higher listening scores, whereas participants with lower vocabulary scores tended to record lower listening scores. Based on the predetermined interpretation criteria, the magnitude of .669 represented a strong association. The 95% confidence interval ranged from .277 to .870 and did not include zero, supporting the conclusion that the population association was positive while also demonstrating considerable uncertainty because of the small sample. The squared correlation coefficient was .447, indicating that the variables shared approximately 44.7% of their variance. Because p was lower than the .05 significance criterion, the null hypothesis of no statistically significant correlation was rejected and the research hypothesis was supported. This finding describes covariation between the measured variables and should not be interpreted as evidence that vocabulary mastery causally determines listening performance.

4 Discussion

The present study found a strong positive relationship between vocabulary mastery and converted TOEFL ITP listening performance, with a correlation of .669 and a significance value of .003; this pattern supports recent evidence that broader and deeper aural lexical knowledge is associated with more successful comprehension of spoken English among university EFL learners (Han & Qian, 2024; Du et al., 2022). The finding means that participants who achieved higher vocabulary scores tended to obtain higher listening scores, but it does not demonstrate that vocabulary alone produced those outcomes; comparable investigations have linked phonological or aural vocabulary measures with listening performance while emphasizing that lexical knowledge operates alongside grammar, processing efficiency, and other comprehension resources (Masrai, 2022; Wang, 2023b). The shared variance of approximately 44.7% further indicates a meaningful overlap between the two constructs, while leaving more than half of the observed listening variance unaccounted for by their bivariate relationship and therefore open to alternative learner, task, and contextual explanations.

One explanation for the association is that successful listening requires learners to retrieve lexical representations from rapidly unfolding acoustic input rather than merely recognize words presented orthographically; research comparing aural and written vocabulary has shown that these knowledge modalities are related but distinct, while aural decoding ability is strongly connected to overall comprehension (Hamada & Yanagawa, 2024; Ke & Wang, 2022). Learners who cannot segment connected speech, distinguish phonologically similar forms, or access meanings quickly may lose subsequent information while attempting to resolve an earlier word, thereby increasing processing demands throughout TOEFL conversations and talks; analyses of perceptual errors and real-time listening problems identify word recognition, speech segmentation, syntactic parsing, and contextual integration as interacting sources of

difficulty (Wong et al., 2021; Goh, 2023). Vocabulary mastery may consequently support TOEFL ITP listening by making lexical access more efficient, yet the strength of this support will depend on whether vocabulary knowledge is sufficiently precise, automatized, and available in the auditory modality.

The descriptive results add a qualification because most participants demonstrated fair vocabulary mastery, whereas most listening scores fell within the poor or very poor categories; this discrepancy suggests that knowledge captured by the written multiple-choice vocabulary test was not fully available during auditory processing, a distinction observed for phrasal verbs and productive lexical knowledge (Kang, 2024; Allal-Sumoto et al., 2023). TOEFL ITP listening further requires comprehension across short conversations, extended exchanges, and academic talks, so learners must recognize both frequent vocabulary and context-specific expressions under time pressure; corpus-based studies of IELTS materials and academic spoken discourse show that adequate comprehension requires substantial lexical coverage and that demands vary across sections, genres, and topics (Phung & Ha, 2022; Trang et al., 2023). Thus, a fair vocabulary score should not be interpreted as sufficient listening readiness, particularly when the vocabulary instrument does not measure pronunciation-sensitive recognition, formulaic sequences, contextual meanings, or the speed with which known words can be retrieved from speech.

The unexplained proportion of listening performance is consistent with multidimensional accounts in which aural vocabulary size contributes to comprehension together with metacognitive knowledge, self-efficacy, attention, background knowledge, and strategic processing; diagnostic comparisons also demonstrate that receptive subskills can function differently across listening and reading modalities and proficiency levels (Du & Man, 2022; He et al., 2022). Individual differences in planning, monitoring, directed attention, inference, and recovery from missed information may therefore help explain why participants with similar vocabulary scores obtained different listening results, properties of the two tests may also have influenced the size of the observed correlation; strategy measurement and vocabulary-test validation research shows that scores depend on construct definition, item selection, word grouping, reference corpora, and response mode (Wakamoto & Rose, 2021; Pinchbeck et al., 2022). The coefficient should accordingly be interpreted as a context-bound estimate of association between two operational measures, not as a complete explanatory model of TOEFL ITP listening performance or a universal vocabulary threshold.

Pedagogically, the findings support integrating vocabulary development with auditory practice instead of treating vocabulary as an exclusively written list-learning activity; captioned input and transcript-supported listening can help learners connect orthographic forms with pronunciations and meanings, although their effectiveness varies with proficiency, question difficulty, and playback speed (Teng, 2022; Mahalingappa et al., 2024). Instruction may gradually move from supported exposure to single-play practice by combining repeated encounters, retrieval of meanings from speech, connected-speech analysis, and comprehension of increasingly extended academic input; comparisons of listening, reading-while-listening, and viewing conditions indicate that multimodal exposure and prior lexical knowledge jointly shape immediate learning and retention of words and collocations (Teng, 2024; Dang et al., 2022a). These implications concern instructional priorities rather than effects demonstrated by the present correlational design, so future classroom interventions should directly test whether strengthening aural vocabulary recognition improves TOEFL ITP listening outcomes beyond ordinary test familiarization and whether any gains remain detectable after textual or replay support is withdrawn.

Several limitations require restrained interpretation: the purposively selected sample contained 17 seventh-semester students from one institution, producing a wide confidence interval for the correlation and restricting generalization beyond the participating cohort; methodological syntheses caution that characterized or narrowly recruited samples can

exaggerate universality and obscure variation across linguistic and sociocultural populations (Plonsky, 2023; Godfroid & Andringa, 2023). Moreover, the vocabulary and listening instruments were adapted from publicly available instructional materials, while evidence concerning reliability, item functioning, criterion validity, and equivalence to an official TOEFL ITP form was not reported; research on EAP vocabulary materials and high-stakes language goals emphasizes that score interpretations must be supported by alignment among test content, target constructs, learner populations, and intended uses (Lu & Dang, 2022; Fung & Chung, 2024). Future research should recruit larger multisite samples, employ validated aural and written vocabulary measures alongside a listening instrument, report reliability and confidence intervals, examine influential cases, and model additional cognitive, linguistic, affective, and experiential predictors.

5 Conclusion

This study demonstrated a strong positive relationship between vocabulary mastery and TOEFL ITP listening performance among seventh-semester Indonesian EFL undergraduates at Cokroaminoto Palopo University. The Pearson correlation analysis produced a coefficient of .669 with a significance value of .003, indicating that participants with higher vocabulary scores tended to achieve higher listening scores. The two variables shared approximately 44.7% of their variance, although the confidence interval indicated uncertainty surrounding the population estimate. Descriptively, vocabulary mastery was predominantly classified as fair, whereas listening performance was concentrated in the poor and very poor categories. This contrast indicates that general vocabulary knowledge may not always be readily accessible during real-time listening. The study contributes context-specific evidence by examining the vocabulary–listening relationship through TOEFL ITP performance among Indonesian students approaching undergraduate program completion, a population and assessment context that remain less represented in recent research than general listening, TOEIC, IELTS, and experimentally designed listening assessments.

These findings highlight the importance of integrating vocabulary development with auditory recognition, connected-speech processing, contextual meaning retrieval, and progressively demanding listening practice. Nevertheless, the results should be interpreted cautiously because the study employed a correlational design, involved only 17 purposively selected students from one institution, and did not establish a causal relationship between vocabulary mastery and listening performance. Furthermore, the instruments were adapted from publicly available instructional materials, while evidence concerning their reliability, item functioning, and equivalence to an official TOEFL ITP form was not reported. Future research should recruit larger and more diverse samples from multiple institutions, employ validated measures of written and aural vocabulary knowledge, use an authorized standardized listening instrument, and examine additional predictors such as phonological decoding, working memory, metacognitive knowledge, anxiety, accent familiarity, and test-taking experience. Experimental or longitudinal studies are also needed to determine whether targeted aural vocabulary instruction produces sustained improvements in TOEFL ITP listening performance.

References

- Allal-Sumoto, T. K., Miyoshi, K., & Mizuhara, H. (2023). The effect of productive vocabulary knowledge on second language comprehension. *Frontiers in Psychology*, 14, Article 1049885. <https://doi.org/10.3389/fpsyg.2023.1049885>

- Chung, E., Wan, A., & Fung, D. (2024). Understanding academic vocabulary learning in higher education: Perspectives from first-year undergraduates in Hong Kong. *International Journal of Applied Linguistics*, 34(4), 1368–1384. <https://doi.org/10.1111/ijal.12576>
- Dang, T. N. Y., Lu, C., & Webb, S. (2022a). Incidental learning of collocations in an academic lecture through different input modes. *Language Learning*, 72(3), 728–764. <https://doi.org/10.1111/lang.12499>
- Dang, T. N. Y., Lu, C., & Webb, S. (2022b). Incidental learning of single words and collocations through viewing an academic lecture. *Studies in Second Language Acquisition*, 44(3), 708–736. <https://doi.org/10.1017/S0272263121000474>
- Du, G., Hasim, Z., & Chew, F. P. (2022). Contribution of English aural vocabulary size levels to L2 listening comprehension. *International Review of Applied Linguistics in Language Teaching*, 60(4), 937–956. <https://doi.org/10.1515/iral-2020-0004>
- Du, G., & Man, D. (2022). Person factors and strategic processing in L2 listening comprehension: Examining the role of vocabulary size, metacognitive knowledge, self-efficacy, and strategy use. *System*, 107, Article 102801. <https://doi.org/10.1016/j.system.2022.102801>
- Fung, D., & Chung, E. (2024). Defining language goals in EMI: Vocabulary demand in a high-stakes assessment in Hong Kong. *Language and Education*, 38(2), 188–202. <https://doi.org/10.1080/09500782.2023.2219654>
- Giordano, M. J. (2023). Lexical coverage in dialogue listening. *Language Teaching Research*, 27(6), 1504–1531. <https://doi.org/10.1177/1362168821989869>
- Godfroid, A., & Andringa, S. (2023). Uncovering sampling biases, advancing inclusivity, and rethinking theoretical accounts in second language acquisition: Introduction to the special issue SLA for all? *Language Learning*, 73(4), 981–1002. <https://doi.org/10.1111/lang.12620>
- Goh, C. C. M. (2023). Learners' cognitive processing problems during comprehension as a basis for L2 listening research. *System*, 119, Article 103164. <https://doi.org/10.1016/j.system.2023.103164>
- Hamada, Y., & Yanagawa, K. (2024). Aural vocabulary, orthographic vocabulary, and listening comprehension. *International Review of Applied Linguistics in Language Teaching*, 62(2), 953–975. <https://doi.org/10.1515/iral-2022-0100>
- Han, D., & Qian, D. D. (2024). Evaluating the roles of breadth and depth of aural vocabulary knowledge in listening comprehension of EFL learners: An investigation applying auditory measures. *System*, 120, Article 103207. <https://doi.org/10.1016/j.system.2023.103207>
- He, L., Xiong, L., & Min, S. (2022). Diagnosing listening and reading skills in the Chinese EFL context: Performance stability and variability across modalities and performance levels. *System*, 106, Article 102787. <https://doi.org/10.1016/j.system.2022.102787>
- Kang, D.-M. (2024). L2 learners' aural and orthographic phrasal verb knowledge in relation to their listening comprehension. *System*, 123, Article 103331. <https://doi.org/10.1016/j.system.2024.103331>
- Ke, Z., & Wang, Y. (2022). Exploring the relationship between aural decoding and listening comprehension among L2 learners of English. *System*, 104, Article 102688. <https://doi.org/10.1016/j.system.2021.102688>
- Lu, C., & Dang, T. N. Y. (2022). Vocabulary in EAP learning materials: What can we learn from teachers, learners, and corpora? *System*, 106, Article 102791. <https://doi.org/10.1016/j.system.2022.102791>

- Mahalingappa, L., Zong, J., & Polat, N. (2024). The impact of captioning and playback speed on listening comprehension of multilingual English learners at varying proficiency levels. *System*, 120, Article 103192. <https://doi.org/10.1016/j.system.2023.103192>
- Masrai, A. (2022). The relationship between two measures of L2 phonological vocabulary knowledge and L2 listening comprehension. *TESOL Journal*, 13(1), Article e612. <https://doi.org/10.1002/tesj.612>
- Phung, D. H., & Ha, H. T. (2022). Vocabulary demands of the IELTS listening test: An in-depth analysis. *SAGE Open*, 12(1), Article 21582440221079934. <https://doi.org/10.1177/21582440221079934>
- Pinchbeck, G. G., Brown, D., McLean, S., & Kramer, B. (2022). Validating word lists that represent learner knowledge in EFL contexts: The impact of the definition of word and the choice of source corpora. *System*, 106, Article 102771. <https://doi.org/10.1016/j.system.2022.102771>
- Plonsky, L. (2023). Sampling and generalizability in Lx research: A second-order synthesis. *Languages*, 8(1), Article 75. <https://doi.org/10.3390/languages8010075>
- Reynolds, B. L., Xie, X. S., & Pham, Q. H. P. (2022). Incidental vocabulary acquisition from listening to English teacher education lectures: A case study from Macau higher education. *Frontiers in Psychology*, 13, Article 993445. <https://doi.org/10.3389/fpsyg.2022.993445>
- Saeedakhtar, A., Haqju, R., & Rouhi, A. (2021). The impact of collaborative listening to podcasts on high school learners' listening comprehension and vocabulary learning. *System*, 101, Article 102588. <https://doi.org/10.1016/j.system.2021.102588>
- Teng, M. F. (2022). Incidental L2 vocabulary learning from viewing captioned videos: Effects of learner-related factors. *System*, 105, Article 102736. <https://doi.org/10.1016/j.system.2022.102736>
- Teng, M. F. (2024). Working memory and prior vocabulary knowledge in incidental vocabulary learning from listening, reading, reading-while-listening, and viewing captioned videos. *System*, 124, Article 103381. <https://doi.org/10.1016/j.system.2024.103381>
- Trang, N. H., Nguyen, D. T. B., & Ha, H. T. (2023). Vocabulary demands of academic spoken English revisited: A case of university lectures and TED presentations. *SAGE Open*, 13(1), Article 21582440231155334. <https://doi.org/10.1177/21582440231155334>
- Wakamoto, N., & Rose, H. (2021). Learning to listen strategically: Developing a listening comprehension strategies questionnaire for learning English as a global language. *System*, 103, Article 102670. <https://doi.org/10.1016/j.system.2021.102670>
- Wang, Y. (2023a). Explaining L2 learners' listening comprehension: The role of aural vocabulary knowledge. *Sintagma*, 35, 103–118. <https://doi.org/10.21001/sintagma.2023.35.07>
- Wang, Y. (2023b). Explaining listening comprehension among L2 learners of English: The contribution of vocabulary and grammar knowledge. *Porta Linguarum Revista Interuniversitaria de Didáctica de las Lenguas Extranjeras*, 39, 63–74. <https://doi.org/10.30827/portalin.vi39.22286>
- Wong, S. W. L., Leung, V. W. H., Tsui, J. K. Y., Dealey, J., & Cheung, A. (2021). Chinese ESL learners' perceptual errors of English connected speech: Insights into listening comprehension. *System*, 98, Article 102480. <https://doi.org/10.1016/j.system.2021.102480>