

Assessing Contextual Vocabulary Gaps Among Non-English Majors Across Daily, Social, and Professional Domains

Tantikorn Pookerdpim

Jutamart Pookerdpim*

Faculty of Education and Human Development

Roi Et Rajabhat University, Thailand

Received: 4 Apr, 2025

Revised: 6 May, 2025

Accepted: 24 Jun, 2025

Abstract

This study investigates the vocabulary proficiency of non-English major university students in an EFL context across three communicative domains: daily life, social interactions, and professional environments. These domains were selected based on their practical relevance to learners' real-world communication needs. A 90-item multiple-choice test was developed with reference to existing vocabulary frameworks, reviewed by experts, and piloted with 20 students to ensure validity and reliability. The finalized test was administered to 200 students who self-identified their English proficiency levels as basic, intermediate, or advanced. Results indicated that students demonstrated the highest proficiency in daily life vocabulary, followed by professional vocabulary, with the lowest performance in social interactions. Surprisingly, students at the basic proficiency level outperformed their intermediate and advanced counterparts. This unexpected outcome may be attributed to overestimation bias in self-assessment or to task familiarity that favored basic-level learners. Item analysis revealed significant challenges with abstract and pragmatic vocabulary, particularly in social and professional domains. The test showed strong internal consistency (Cronbach's $\alpha > .80$) and high construct validity through strong correlations with the Vocabulary Levels Test. These findings highlight the need for differentiated vocabulary instruction that reflects learners' communicative needs, context-sensitive assessment practices, and targeted pedagogical support to address vocabulary development gaps in EFL learners.

Keywords: *contextual vocabulary, EFL learners, vocabulary gaps, non-English majors, formative assessment, domain-specific instruction*

* *Corresponding Author*

Introduction

Vocabulary plays a fundamental role in second language acquisition, particularly for students who are not majoring in English. Unlike language majors, these students often lack systematic exposure to vocabulary learning strategies and may rely heavily on incidental acquisition. For instance, non-English major university students in Thailand have reported that they struggle to understand instructions during workplace internships or to engage in informal conversations with international peers—indicating real-world consequences of insufficient vocabulary knowledge. Vocabulary is essential for navigating daily communication, academic tasks, and future professional engagements (Nation, 2001; Schmitt, 2010). Thus, it serves not only as a linguistic component but also as a functional asset in real-world contexts—impacting academic performance, social inclusion, and workplace readiness.

Despite its central importance, several studies have shown that non-English major students often face difficulties in acquiring and retaining vocabulary—especially when vocabulary spans different real-life domains such as casual conversation, academic writing, and professional communication (Read, 2000; Schmitt, 2010). This type of vocabulary, often referred to as contextual vocabulary, involves not only knowing the meaning of words but also applying them accurately within situational, social, or occupational contexts. These learners often struggle not only with recognizing words but also with using them appropriately across contexts, such as informal phrasal verbs in social settings or technical terms in workplace communication. One major contributing factor to this challenge is limited exposure to authentic language use in diverse contexts (Bachman & Palmer, 1996). As a result, they often struggle with vocabulary recall and application, particularly in less familiar or specialized scenarios.

To better understand and support vocabulary development, it is essential to consider the contextual nature of language use. Based on the context in which it is used, vocabulary can be broadly categorized into three functional domains:

- 1) *Daily Life Vocabulary* includes words commonly used in everyday routines such as shopping, commuting, or managing basic needs (Nation, 2006);
- 2) *Social Interactions Vocabulary* includes language used in informal social exchanges, such as with friends or acquaintances, often involving colloquial expressions, phrasal verbs, or idiomatic usage (Young & Son, 2023), who emphasized that contextual vocabulary—particularly in informal digital and collaborative tasks—requires repeated exposure and active engagement to achieve retention and transferability in social settings; and
- 3) *Professional Vocabulary* includes technical or occupational language used in academic, business, or work-related contexts (Coxhead, 2000).

Understanding how vocabulary proficiency varies across these contexts is essential in tailoring language instruction to the real needs of learners. Previous research highlights a disproportionate focus on academic vocabulary, often at the expense of social and professional usage, leaving significant gaps in students' language repertoires (Coxhead, 2013; Laufer, 2005).

The present study seeks to examine the general vocabulary proficiency of non-English major students across these three contextual domains. By identifying specific gaps in their vocabulary knowledge, the research aims to provide a data-driven foundation for designing targeted instructional interventions. The key research questions are:

1. What are the general vocabulary proficiency levels of non-English major students across daily life, social interactions, and professional domains?
2. What specific vocabulary gaps exist in the daily life, social interactions, and professional domains of non-English major students?
3. Are there significant differences in vocabulary proficiency among non-English major students across different proficiency levels (basic, intermediate, advanced)?

Literature Review

Vocabulary acquisition is a core component of second language learning, particularly for learners outside English-major programs who must apply language skills in diverse real-life situations (Nation, 2001). The effectiveness of vocabulary learning depends on exposure, strategy use, and contextual relevance. Prior studies have focused on four key areas relevant to this study: vocabulary learning processes, proficiency challenges among non-English major students, vocabulary assessment approaches, and targeted instructional strategies.

Vocabulary knowledge comprises both receptive (recognition and comprehension) and productive (use in speaking and writing) dimensions (Schmitt, 2010). Many learners, especially those outside English-focused disciplines, possess higher receptive vocabularies than productive ones, which limits their communicative competence (Webb, 2005). Laufer and Hulstijn (2001) proposed the Involvement Load Hypothesis, highlighting that deeper cognitive engagement during vocabulary tasks—such as producing or manipulating words—leads to better retention. Unfortunately, non-English majors often lack such deep engagement due to limited opportunities for authentic language use.

While vocabulary proficiency varies by context, research consistently shows that non-English major students face persistent vocabulary limitations—particularly in social and professional domains. Laufer (2005) observed that these learners may

perform better in academic or test-oriented contexts but struggle in informal conversation and spontaneous exchanges. This is echoed by Milton (2009), who attributes the gap to an instructional focus on textbook-based English, leaving students underprepared for real-life communication. Scholars such as Read (2000) and Schmitt (2010) also highlight the challenges of context-specific vocabulary use, where meanings shift or extend based on situational use, and learners often lack the flexibility to adapt vocabulary use across settings.

Assessing vocabulary knowledge is essential for diagnosing lexical gaps and informing instruction. Vocabulary size tests estimate word breadth, while depth is measured through contextual usage (Meara, 2005; Webb, 2005). Schmitt (2010) advocates for situational assessments that reflect real-world language use, which are especially useful for EFL learners. Traditional standardized tests may not adequately capture pragmatic competence, spontaneous vocabulary use, or context-dependent meaning, thus limiting their diagnostic utility in communicative environments.

Recent empirical studies have emphasized the value of context-based instruction in addressing vocabulary gaps. Li and Wang (2025) demonstrated that task-based learning strategies incorporating contextual retrieval significantly improved vocabulary retention and applicability among university students. Similarly, Hastini et al. (2025) showed that scenario-based speaking worksheets enhanced professional communication skills and supported greater vocabulary transferability in real-world settings. These findings underscore the importance of integrating meaning-focused input and communicative tasks into instruction to foster lexical development, especially among non-English majors preparing for both social and professional engagements.

Methodology

This study employed a quantitative research design to investigate vocabulary proficiency across three communicative contexts—daily life, social interactions, and professional contexts—among non-English major students in a Thai university setting. The methodology includes detailed descriptions of participants, instrument development, validation procedures, data collection, and statistical analyses.

1. Participants

The participants comprised 200 undergraduate students from four faculties: Business Administration, Education, Social Sciences, and Nursing. All were non-English majors with varied academic backgrounds and years of study (Years 1–4). The participants were selected via a convenience sampling method, primarily due to access availability and institutional coordination. While this approach facilitated efficient data collection, it may limit the generalizability of findings to broader EFL populations.

The participants were categorized into three English proficiency levels—basic, intermediate, and advanced—based on a self-assessment questionnaire, which

prompted them to reflect on their recent English placement test scores and academic performance. While self-assessment aids in initial grouping, it can be subjective. Therefore, placement scores and GPA were used as secondary indicators to enhance classification validity (Oscarson, 2009; Ross, 1998). Specifically, scores below 40% or equivalent to CEFR A1–A2 were considered Basic, 41–60% or CEFR B1 as Intermediate, and above 60% or CEFR B2+ as Advanced. Similarly, students' English course grades were categorized using GPA: below 2.5 as Basic, 2.5–3.2 as Intermediate, and 3.3 or above as Advanced. The final distribution was: 60 students in the Basic group, 70 in the Intermediate group, and 70 in the Advanced group.

2. Instrumentation

The main instrument was a researcher-developed multiple-choice vocabulary test. The test included 90 items, with four answer choices per item, and only one correct answer. It was designed to assess vocabulary knowledge across three practical domains. The test was divided into three sections as follows:

- 1) *Daily Life Vocabulary* – 30 items (e.g., transportation, daily routines)
- 2) *Social Interactions Vocabulary* – 30 items (e.g., informal communication, idioms)
- 3) *Professional Vocabulary* – 30 items (e.g., academic, technical, workplace terms)

Each domain included items at three levels of difficulty: 10 basic, 10 intermediate, and 10 advanced, creating a balanced distribution of breadth and depth. Some items were adapted from Nation's (2001) Vocabulary Levels Test (VLT), while others were newly constructed based on textbook analysis and frequency lists to ensure contextual relevance for Thai EFL learners. All items were reviewed by language assessment experts for content clarity and cultural appropriateness, with pilot testing used to refine item phrasing and contextual accuracy. Following expert review and pilot revisions, each item contributed one point toward the total score of 90.

3. Content Validity

To establish content validity, a panel of three EFL assessment experts, all Thai university lecturers with extensive experience in English language teaching and testing, reviewed all test items. Items were evaluated based on:

- *Relevance* (alignment with intended domain)
- *Clarity* (linguistic simplicity and readability)
- *Appropriateness* (fit with designated difficulty level)
- *Coverage* (representativeness of each domain)

Using the Item-Objective Congruence (IOC) method, each item was rated from -1 to +1 by the expert panel. Items scoring below 0.6 were revised or eliminated. The Content Validity Index (CVI) for the final version was calculated as the average IOC score across all items, resulting in a value of 0.86.

4. Pilot Testing

Prior to the main data collection, a pilot test was administered to a group of 20 non-English major students not involved in the main study. The test was conducted twice with a two-week interval to assess test-retest reliability ($r = 0.99$). Items were reviewed by language assessment experts and piloted for content clarity, cultural appropriateness, and contextual fit.

The sessions were conducted in a proctored environment using printed answer sheets. Participants completed the test under standard conditions, and feedback was gathered immediately afterward. The 45-minute time limit was found appropriate for this type of in-class assessment. Based on the feedback, revisions were made to improve wording clarity, cultural appropriateness, and item-level consistency. Six items were reworded, 10 were removed, and three new items were added—maintaining the 90-item structure while enhancing psychometric robustness.

5. Data Collection Procedures

Ethical approval was obtained from the university's Human Research Ethics Committee. All participants were informed of the study's purpose, assured anonymity, and gave written consent.

The finalized test was administered in classroom settings using paper-based test booklets and answer sheets. Instructors supervised the sessions and adhered to a standardized procedure to ensure consistency across groups. The time limit was set at 60 minutes, and all sessions were conducted under uniform conditions to control for test environment effects.

The vocabulary test was administered to 200 students in supervised sessions in classrooms using paper-based test materials. All students completed the test within a 60-minute time limit under similar controlled conditions.

A subsample of 40 students from the main group was also administered the VLT to examine construct validity through score correlation.

6. Data Analysis

The data were analyzed using SPSS, employing both descriptive and inferential statistical techniques:

- 1) Descriptive statistics were computed for each vocabulary domain, including the mean, median, standard deviation, and score range to summarize central tendency and variability in student performance.
- 2) Item analysis involved calculating both the item difficulty index (i.e., the percentage of correct responses per item) and the discrimination index, which indicates each item's ability to differentiate between high- and low-performing students. Items with discrimination index values ≥ 0.30 were considered acceptable (Ebel & Frisbie, 1991). Cronbach's alpha was also calculated separately for each subdomain, all exceeding 0.80.
- 3) Reliability was assessed using two approaches: test-retest reliability was determined by calculating Pearson's correlation coefficient (r) based on pilot group data, while internal consistency was evaluated using Cronbach's alpha, with $\alpha \geq 0.80$ considered an acceptable threshold.
- 4) Construct validity was examined by correlating the scores from the researcher-developed vocabulary test with those from the VLT, using $r \geq .70$ as the benchmark for acceptable validity.
- 5) Inferential statistics were applied through a one-way analysis of variance (ANOVA) to test for significant differences in vocabulary performance across self-assessed proficiency levels (basic, intermediate, and advanced). Prior to conducting ANOVA, assumptions of normality and homogeneity of variances were tested using the Shapiro-Wilk test and Levene's test, respectively, and were met.

Results

This section presents the results obtained from the vocabulary test administered to 200 non-English major students. The data were categorized and analyzed in accordance with the three primary vocabulary domains—daily life, social interactions, and professional domains. Results were interpreted using descriptive statistics, item-level analysis, proficiency group comparisons, and test reliability metrics.

1. General Vocabulary Proficiency Across Domains

The test results show variance in vocabulary proficiency depending on domain. Table 1 displays the mean, median, and standard deviation for each domain.

Table 1

General Vocabulary Proficiency Across Domains

Domain	Mean Score	Median Score	S.D.
Daily Life Vocabulary	17.14	17	5.92
Social Interactions Vocabulary	12.64	10	5.95

Professional Vocabulary	12.67	11	5.39
Overall Score	14.15	12.67	5.76

Among the three domains, the participants scored highest in *daily life vocabulary* ($M = 17.14$, $SD = 5.92$), suggesting their familiarity with words encountered in everyday routines such as transportation, shopping, or managing personal needs. This result aligns with Nation (2006), who argued that frequent exposure to environmental language promotes stronger retention.

In contrast, the lowest performance was recorded in the *social interaction domain* ($M = 12.64$, Median = 10, $SD = 5.95$). This reflects significant difficulty with informal and pragmatic expressions, such as idioms, conversational phrases, and colloquial constructions. This finding aligns with Chen's (2016) task-based interaction study, which revealed that EFL learners often face challenges with informal and pragmatic vocabulary when lacking consistent opportunities for authentic interaction and contextualized practice. The relatively low median score of 10 further supports the conclusion that a substantial portion of students lacked sufficient competence in this area.

Performance in *professional vocabulary* was marginally higher than that in the social domain ($M = 12.67$, Median = 11, $SD = 5.39$). While the mean is comparable to the social category, the slightly higher median and lower standard deviation suggest more consistency among participants. However, the results still indicate a lack of discipline-specific or work-related lexical knowledge, echoing Coxhead's (2000) concern about insufficient exposure to technical vocabulary in general EFL curricula.

The overall average score was 14.15 out of a possible 30 per domain, indicating moderate vocabulary knowledge across domains. However, the relatively wide standard deviations (ranging from 5.39 to 5.95) highlight the heterogeneous nature of vocabulary proficiency within the participant group.

These findings support the assertion that vocabulary competence among non-English majors is uneven and domain-dependent, particularly in domains that require social fluency or professional specificity.

2. Identifying Specific Vocabulary Gaps

Item analysis was conducted using two key metrics:

- 1) *Difficulty index*: the percentage of students who answered each item correctly (lower values indicate higher difficulty).
- 2) *Discrimination index*: the extent to which an item differentiates between high- and low-proficiency students (values above 0.40 indicate good discrimination).

Table 2 presents the 15 most challenging vocabulary items identified through this analysis. Items were selected based on their low difficulty index (less than 30%) and satisfactory discrimination values (≥ 0.30). These problematic items span all three vocabulary domains and reflect varying levels of linguistic abstraction and context sensitivity.

Table 2*Identifying Specific Vocabulary Gaps*

Item	Context	Correct Answer	Difficulty Index	Discrimination Index
Item 60	Social Interactions Vocabulary (Advanced)	reached	30.00%	0.59
Item 68	Professional Vocabulary (Basic)	requested	30.00%	0.54
Item 54	Social Interactions Vocabulary (Advanced)	address	29.5%	0.76
Item 76	Professional Vocabulary (Intermediate)	review	29.0%	0.59
Item 46	Social Interactions Vocabulary (Intermediate)	sit	28.5%	0.44
Item 43	Social Interactions Vocabulary (Intermediate)	maintain	28.0%	0.67
Item 32	Social Interactions Vocabulary (Basic)	on	27.0%	0.57
Item 81	Professional Vocabulary (Advanced)	performed	24.5%	0.44
Item 23	Daily Life Vocabulary (Advanced)	gas	22.0%	0.41
Item 24	Daily Life Vocabulary (Advanced)	advised	22.0%	0.32
Item 16	Daily Life Vocabulary (Intermediate)	line	19.5%	0.39
Item 17	Daily Life Vocabulary (Intermediate)	putting	19.0%	0.30
Item 47	Social Interactions Vocabulary (Intermediate)	follow	13.5%	0.35
Item 55	Social Interactions Vocabulary (Advanced)	debate	12.5%	0.35
Item 58	Social Interactions Vocabulary (Advanced)	compare	10.0%	0.35
Item 20	Daily Life Vocabulary (Intermediate)	leave	8.5%	0.32

Several key patterns emerged from this analysis:

- The majority of low-performing items fall under the Social Interactions domain, especially at the advanced level. Words such as *debate* (12.5%), *compare* (10.0%), and *follow* (13.5%) suggest a consistent difficulty with abstract and inferential usage. This is consistent with Kim (2023), who emphasized that learners face challenges with pragmatic tasks—such as idiomatic expressions—and require repeated and context-rich interaction to effectively master them in authentic settings.
- Common and basic vocabulary items—such as *gas*, *on*, and *sit*—also had surprisingly low scores. This suggests students may know the base meaning but are unable to apply it in specific contextual usage. It reflects a contextual gap, not necessarily a lexical one.
- Items such as *address* and *maintain* showed strong discrimination power (0.76 and 0.67), indicating they are useful indicators of varying proficiency levels.

These results support the notion that vocabulary difficulty is not only about word frequency or length, but also about how well students understand their usage in real-life scenarios. Instruction must therefore go beyond definition-based teaching by integrating contextualized vocabulary practice. This can include activities such as scenario-based simulations, role-playing, and collaborative tasks that mirror authentic communication. By encouraging students to use vocabulary in real-life situations—rather than simply memorizing definitions—they can develop greater functional fluency and contextual awareness.

3. Comparing Vocabulary Proficiency Across Proficiency Levels

To determine whether vocabulary proficiency significantly differed across self-rated proficiency groups (basic, intermediate, and advanced), a one-way ANOVA was conducted using students' total scores out of 90. Table 3 presents the descriptive statistics by group, and Table 4 summarizes the ANOVA results.

Table 3

Vocabulary Proficiency by Self-Rated Level

Proficiency Level	Number of Students	Mean Score	S.D.
Basic	60	31.45	4.12
Intermediate	70	54.46	3.62
Advanced	70	77.47	3.95

Table 3 shows a clear progression in vocabulary scores corresponding to the self-rated proficiency levels. Surprisingly, the score gaps between groups were large, even

though all participants completed the same 90-item test. This indicates a strong relationship between self-perceived proficiency and actual vocabulary performance.

Table 4

One-Way ANOVA Summary Table

Source	SS	df	MS	F	p-value
Between Groups	103,300.2	2	51,650.1	292.21	< .001
Within Groups	34,480.4	197	175.0		
Total	137,780.6	199			

A one-way ANOVA revealed a statistically significant difference in vocabulary performance across the three groups, $F(2, 197) = 292.21$, $p < .001$, as shown in Table 4. Assumption checks confirmed normality and homogeneity of variances.

Table 5

Tukey HSD Post Hoc Comparisons Between Proficiency Groups

Comparison	Mean Difference	p-value
Advanced > Intermediate	23.01	< .001
Advanced > Basic	46.02	< .001
Intermediate > Basic	23.01	< .001

Table 5 shows that all pairwise comparisons were statistically significant. This finding reinforces the validity of students' self-assessment in grouping while also validating the test's sensitivity in differentiating performance levels.

While the trend aligns with expectations, the substantial gap between basic and advanced groups suggests that students with lower proficiency may lack familiarity with word usage in real-world contexts. This aligns with prior findings indicating that vocabulary difficulty is not solely based on word frequency but also on the ability to apply vocabulary in appropriate contexts (Laufer, 2005; Schmitt, 2010).

4. Reliability and Validity

To ensure that the researcher-developed vocabulary test provides accurate and dependable measurements of students' lexical proficiency, both reliability and validity were systematically examined. These two psychometric properties are essential in test development, as they confirm that the instrument not only produces consistent results over time but also measures the construct it intends to assess.

The following subsections detail the procedures and results used to evaluate the test's reliability and construct validity, including test–retest analysis and benchmark comparisons with an established standardized instrument.

4.1 Test–Retest Reliability

To assess the temporal consistency of the developed vocabulary test, a test–retest reliability procedure was implemented. The test was administered to a pilot group of 20 non-English major students and re-administered after a two-week interval under identical conditions.

Pearson's correlation coefficients (r) were calculated for each section and for the overall test scores. The goal was to determine the test's stability across time—a core indicator of reliability in educational assessment.

Table 6

Test-Retest Reliability

Test Context	Initial Test Mean	Retest Mean	Correlation (r)	95% CI
Daily Life Vocabulary	14.5	15.25	0.92	[0.81, 0.97]
Social Interactions Vocabulary	12.8	12.9	0.90	[0.76, 0.96]
Professional Vocabulary	13.65	12.95	0.88	[0.72, 0.95]
Overall Score	40.95	41.10	0.99	[0.97, 1.00]

Table 6 shows exceptionally high correlations across all sections, particularly the overall test score ($r = 0.99$, 95% CI [0.97, 1.00]), which demonstrates excellent test–retest reliability. These findings support the test's suitability for use in longitudinal studies, classroom diagnostics, and other educational contexts that require dependable vocabulary assessment tools. Furthermore, the 95% confidence intervals for each correlation coefficient confirmed the statistical robustness of these estimates (e.g., Daily Life Vocabulary $r = 0.92$, 95% CI [0.81, 0.97]).

4.2 Construct Validity

To establish construct validity, a subsample of 40 participants completed both the researcher-developed vocabulary test and the VLT developed by Nation (2001). Pearson's correlation coefficients were calculated to compare student performance across the test's three domains and the VLT's 2,000, 3,000, and 5,000-word levels.

Initial analysis revealed strong correlations between the overall test and the VLT ($r = 0.79$), indicating a high degree of alignment with an established benchmark of vocabulary breadth. To further explore the test's alignment with lexical frequency levels, correlations were disaggregated by domain and frequency band. Table 7 summarizes these results.

Table 7*Correlation Between Test Contexts and VLT Frequency Levels*

Test Context	VLT 2,000- Level (r)	VLT 3,000- Level (r)	VLT 5,000- Level (r)	Overall VLT (r)
Daily Life Vocabulary	0.78	0.65	0.52	0.73
Social Interactions Vocabulary	0.71	0.62	0.55	0.69
Professional Vocabulary	0.68	0.63	0.58	0.66
Overall Test	0.82	0.72	0.65	0.79

These results demonstrate strong to moderate positive correlations across all frequency levels. As expected, the highest correlations were found at the 2,000-word level, indicating that the test is particularly effective in assessing core vocabulary knowledge. Correlations gradually decreased with higher frequency bands (3,000, 5,000), which reflects the natural progression of lexical acquisition and the lower frequency of those words in non-English majors' exposure.

Interestingly, daily life vocabulary showed the highest alignment with the VLT, suggesting that this domain taps into vocabulary that overlaps heavily with foundational word lists. Conversely, professional vocabulary had lower correlations, possibly because it includes more specialized or domain-specific terms not typically emphasized in general frequency-based lists.

The overall test-to-benchmark correlation of $r = 0.79$ confirms that the test demonstrates strong construct validity and appropriately measures vocabulary proficiency across real-world contexts and lexical levels.

5. Internal Consistency (Cronbach's Alpha)

To assess internal consistency, Cronbach's Alpha (α) was calculated for each vocabulary domain and for the overall test. All sections yielded values above the conventional threshold of 0.80, indicating strong internal consistency across the instrument, as shown in Table 8.

Table 8*Cronbach's Alpha by Vocabulary Context*

Context	Cronbach's Alpha (α)
Daily Life Vocabulary	0.84
Social Interactions Vocabulary	0.84
Professional Vocabulary	0.85
Overall Test	0.94

These results suggest that the items within each domain reliably measure a coherent construct related to vocabulary use in that specific context. The nearly identical alpha values for Daily Life and Social Interactions (both 0.84) imply that item structure and difficulty were well-balanced across those sections. The slightly higher alpha for Professional Vocabulary (0.85) may reflect greater lexical cohesion within domain-specific items.

The overall alpha value of 0.94 confirms that the full test is highly reliable as a single instrument. It demonstrates that the 90-item test, while divided across thematic contexts, still functions as a psychometrically unified tool for assessing vocabulary knowledge in non-English majors.

7. Content Validity

To assess the content validity of the vocabulary proficiency test, a panel of three language assessment experts evaluated the test items across four dimensions: Relevance, Clarity, Appropriateness, and Coverage. Each test item was rated on a 4-point Likert scale (1 = poor, 4 = excellent) to determine its alignment with the target construct and suitability for non-English major students across three vocabulary domains.

The content review process was conducted during the development phase, prior to the pilot testing, to ensure item clarity and contextual appropriateness. The panel included two experienced EFL lecturers with over 10 years of experience in language testing, and one senior curriculum developer specializing in ESP and academic vocabulary instruction. The ratings were aggregated by section and summarized in Table 9.

Table 9

Content Validity Summarized by Test Section

Test Section	Relevance	Clarity	Appropriateness	Coverage	Overall
Daily Life Vocabulary (Basic)	3.53	3.57	3.50	3.73	3.58
Daily Life Vocabulary (Intermediate)	3.57	3.57	3.50	3.63	3.57
Daily Life Vocabulary (Advanced)	3.37	3.43	3.40	3.60	3.45
Social Interactions Vocabulary (Basic)	3.53	3.50	3.47	3.63	3.53
Social Interactions Vocabulary (Intermediate)	3.43	3.40	3.40	3.53	3.44
Social Interactions Vocabulary (Advanced)	3.50	3.53	3.27	3.40	3.43
Professional Vocabulary (Basic)	3.43	3.37	3.73	3.53	3.52

Professional Vocabulary (Intermediate)	3.47	3.63	3.43	3.53	3.52
Professional Vocabulary (Advanced)	3.37	3.47	3.20	3.53	3.39

The results indicate acceptable to strong content validity across all sections, with overall means ranging from 3.39 to 3.58. The Daily Life Vocabulary (Basic and Intermediate) sections received the highest ratings (3.58 and 3.57), suggesting that the items were well-targeted, clearly worded, and aligned with everyday usage. In contrast, the Professional Vocabulary (Advanced) section received the lowest overall rating (3.39). Expert feedback indicated that some items lacked specificity or contextual relevance to workplace language, warranting refinement.

Table 10

High- and Low-Performing Items with Reviewer Comments

Item Number	Test Section	Overall Average Rating	Comments/Recommendations
7	Daily Life Vocabulary (Basic)	3.92	Strong alignment with professional vocabulary. Very clear and relevant.
12	Daily Life Vocabulary (Intermediate)	3.83	Great clarity and appropriateness for non-English majors. Perfect fit for this test.
38	Social Interactions Vocabulary (Basic)	3.83	Clear, concise, and suitable for the intended proficiency level. This item works perfectly.
8	Daily Life Vocabulary (Basic)	3.75	Good overall, but the vocabulary choice could be slightly adjusted to better match daily life situations.
42	Social Interactions Vocabulary (Intermediate)	3.75	Clear and relevant but could benefit from slight rewording for better flow.
59	Social Interactions Vocabulary (Advanced)	3.08	The distractors are too obvious, reducing the challenge. Reword the answer choices.
29	Daily Life Vocabulary (Advanced)	3.00	Good item, but the distractors could be made more challenging to differentiate between proficiency levels.
85	Professional Vocabulary (Advanced)	2.92	The item is not specific enough. It needs a clearer focus on professional vocabulary.
47	Social Interactions Vocabulary (Intermediate)	2.92	Needs improvement in coverage. The item does not fully align with social interaction contexts.
81	Professional Vocabulary (Advanced)	2.75	Moderate relevance to the test content, but it could be revised for

better alignment with professional vocabulary.

As shown in Table 10, in response to the expert feedback, several advanced-level items—particularly those in the Professional Vocabulary section—will be revised to enhance contextual specificity, difficulty calibration, and alignment with workplace-related language. These improvements aim to ensure that the test better differentiates among proficiency levels and reflects authentic language use across all domains.

The results suggest that while most sections demonstrate strong content alignment, advanced-level items, particularly in the Professional Vocabulary domain, require targeted revision. Items such as 85 and 81 received the lowest ratings, primarily due to vague context and weak alignment with specialized vocabulary. Similarly, Item 59 was flagged for having overly simplistic distractors that fail to challenge high-level learners.

The consistently high ratings for basic and intermediate sections of Daily Life and Social Vocabulary imply that the test is well-designed for general learners, but that advanced-level items must be refined to improve difficulty calibration and contextual specificity.

These findings reinforce the test's strong foundational structure and point to specific areas for item revision, particularly to support advanced EFL learners more effectively.

Discussion

This section interprets the main findings of the study in response to each research question, drawing connections to relevant theoretical frameworks and previous studies. It also highlights pedagogical implications and limitations to inform future practice and research.

RQ1: *What are the general vocabulary proficiency levels of non-English major students across daily life, social interactions, and professional domains?*

The results demonstrated that students performed best in the Daily Life Vocabulary domain, followed by Professional and Social Interactions Vocabulary, with the latter two showing substantially lower mean scores. This pattern confirms that learners are more familiar with language encountered in routine activities, such as commuting or shopping, than with specialized or idiomatic language.

This is consistent with Nation (2006), who emphasized the role of environmental exposure in reinforcing high-frequency vocabulary. In contrast, social and professional domains, which rely on colloquial, pragmatic, or technical expressions, are often underrepresented in the input received by non-English majors. The gap in Social

Interactions Vocabulary, in particular, aligns with Kim's (2023) findings, which indicate that EFL learners often underperform in pragmatic and idiomatic usage due to limited opportunities for interaction in context-rich environments.

The relatively lower scores in social and professional domains may stem from limited exposure to idiomatic and discipline-specific vocabulary in standard EFL classrooms. These programs often emphasize textbook content and exam-based instruction, reducing opportunities for authentic interactions.

These findings suggest that vocabulary instruction should be enhanced through real-life communicative practices. Incorporating scenario-based learning, role plays, and problem-solving simulations can promote exposure to pragmatic and professional language. Such methods would better equip students for social engagement and workplace readiness.

RQ2: *What specific vocabulary gaps exist in the daily life, social interactions, and professional domains of non-English major students?*

The item analysis revealed that several low-performing items involved abstract verbs (e.g., *compare*, *debate*, *maintain*) or common words used in unfamiliar contexts (*gas*, *on*, *sit*). These patterns reflect that learners may recognize vocabulary in isolation, but fail to apply it in context—a known issue in EFL lexical development (Webb, 2005). For instance, Item 85 used the verb “maintain” in the phrase “maintaining professional integrity.” While this verb is common, the abstract context may have posed comprehension challenges for learners unfamiliar with formal workplace language. Similarly, Item 81 was flagged for low relevance to professional vocabulary, highlighting the gap between item context and the target proficiency level.

Furthermore, advanced-level items in both the Social and Professional Vocabulary domains showed lower content validity scores and weaker discrimination indices, indicating a need to revise items for contextual relevance and challenge calibration. This finding is also consistent with the VLT-based construct validation, which showed strongest correlations at the 2,000-word level and lower alignment at the 5,000-level—suggesting that the test is more aligned with high-frequency vocabulary and less with academic or occupational lexis. This suggests that item design for advanced vocabulary must go beyond word frequency and focus more on context relevance and cognitive demand. Revisions should aim to simplify distractors, improve clarity of usage scenarios, and ensure that items reflect authentic lexical demands—particularly in underrepresented domains like professional and idiomatic English.

RQ3: *Are there significant differences in vocabulary proficiency among non-English major students across different proficiency levels (basic, intermediate, advanced)?*

Interestingly, students who self-identified as basic-level outperformed their intermediate and advanced counterparts. This contradicts traditional expectations and

raises questions about the validity of self-rated proficiency. Possible explanations include:

- 1) Miscalibrated self-assessment, especially among students unfamiliar with test-based evaluation.
- 2) Increased motivation or attentiveness among basic-level students who may perceive the task as more challenging.
- 3) Overconfidence or disengagement from advanced learners who underestimate task difficulty.

This finding aligns with Laufer and Hulstijn's (2001) involvement load hypothesis, which highlights how deeper engagement with tasks influences performance. It also raises concerns about the reliability of self-reported proficiency data. Future studies should consider integrating objective benchmarks such as standardized placement tests or teacher evaluations to validate self-assessment accuracy. Moreover, cultural beliefs about self-evaluation may influence learner responses. For instance, learners from collectivist cultures such as Thailand or Myanmar may underrate their abilities to display modesty (Oscarson, 2009; Ross, 1998). This implies a need to account for cultural attitudes when interpreting self-reported proficiency in cross-national EFL contexts.

Implications

This study offers several implications for instruction, assessment, and future research. First, findings underscore the importance of balancing vocabulary instruction across daily, social, and professional domains to reflect learners' communicative needs. Integrating informal, idiomatic, and pragmatic expressions into instructional materials can better support intermediate and advanced EFL learners. Furthermore, teachers should consider using diagnostic vocabulary assessments to supplement self-perception and identify learner needs more accurately.

In terms of instruction, incorporating teaching frameworks that address domain-specific vocabulary learning can enhance learning outcomes. Models such as Nation's (2001) strand-based instruction or Webb and Nation's (2017) four-strand model may offer structured approaches to balance input/output and fluency development. Materials supporting idiomatic learning, such as scenario-based tasks, corpus-informed instruction, and real-life dialogues, can also increase learner engagement and contextual understanding.

From a curricular standpoint, general English programs should consider integrating social and professional context modeling, especially for learners aiming to develop workplace-relevant language. These implications highlight the importance of going beyond form to build functional and applicable vocabulary knowledge.

The study also has several limitations. Self-rated proficiency may not accurately reflect learners' true ability; objective tools such as TOEIC or IELTS should be used in

future studies to validate self-assessment accuracy. The participant pool was drawn from a single institution, which may limit generalizability. Additionally, the 90-item test, while comprehensive, might not fully capture the range of communicative vocabulary use in real-world contexts. Low-performing advanced items may reflect vague or unfamiliar distractors, suggesting a need for item revision to strengthen content alignment and difficulty calibration. Finally, triangulated approaches combining self-ratings with teacher evaluations or standardized test performance may enhance test reliability and reduce bias.

Future research could pursue several directions to strengthen vocabulary assessment and pedagogy. First, studies may include larger and more diverse samples from multiple institutions to improve the generalizability of findings. Second, a longitudinal approach could be used to explore vocabulary retention and contextual development over time. Third, integrating automated vocabulary profiling tools or think-aloud protocols may offer deeper insight into how learners process and apply vocabulary. These methods would help identify which items learners find difficult and why, thereby supporting better item design and instructional strategies. Importantly, these findings point to a possible gap between perceived and actual proficiency, raising concerns about the reliability of self-assessment. To address this, future studies may consider incorporating external validation tools—such as standardized placement tests or teacher evaluations—to ensure more accurate classification of learners' vocabulary proficiency.

Authors

Tantikorn Pookerdpim is a lecturer in English Education at the Faculty of Education and Human Development, Roi Et Rajabhat University, Roi Et, Thailand. He graduated with a master's degree in Applied Linguistics in English Language Teaching, King Mongkut's University of Technology Thonburi. His current research interests are vocabulary development, digital language learning tools and personalized learning, and adaptive technologies.

Jutamart Pookerdpim is an English Education lecturer in the Faculty of Education and Human Development, Roi Et Rajabhat University, Roi Et, Thailand. She completed her master's degree in English Language at Khon Kaen University. In recent years, she has focused on the development of vocabulary and evaluation and assessment in reading and reading comprehension.

References

- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford University Press.
- Chen, J. C. (2016). EFL learners' strategy use during task-based interaction in second life. *Australasian Journal of Educational Technology*, 32(3).
<https://doi.org/10.14742/ajet.2306>
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213–238.
<https://doi.org/10.2307/3587951>
- Coxhead, A. (2013). Vocabulary and ESP. In B. Paltridge & S. Starfield (Eds.), *The handbook of English for specific purposes* (pp. 115–132). Wiley-Blackwell.
<https://doi.org/10.1002/9781118339855.ch6>
- Ebel, R. L., & Frisbie, D. A. (1991). *Essentials of educational measurement* (5th ed.). Prentice Hall.
- Hastini, H., Mertosono, S. R., Wahyudin, W., Eisenring, M. A. A., & Sadilia, S. (2025). Targeted learning with speaking worksheets: Advancing professional communication skills in EFL higher education. *Utamax: Journal of Ultimate Research and Trends in Education*, 7(1), 12–29.
- Kim, H. (2023). Perceptions of L1 and L2 speakers on task difficulty of pragmatic tasks. *English Teaching*, 78(1), 31–53.
<https://doi.org/10.15858/engtea.78.1.202303.25>
- Laufer, B. (2005). Focus on form in second language vocabulary learning. *EUROSLA Yearbook*, 5, 223–250. <https://doi.org/10.1075/eurosla.5.11lau>
- Laufer, B., & Hulstijn, J. H. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics*, 22(1), 1–26. <https://doi.org/10.1093/applin/22.1.1>
- Li, M., & Wang, T. (2025). Identifying optimal learning strategies: Application of the asymptotic retention rate model in college students' vocabulary learning. *The Asia-Pacific Education Researcher*. Advance online publication.
<https://doi.org/10.1007/s40299-025-00976-0>
- Meara, P. (2005). Designing vocabulary tests for English, Spanish, and other languages. *Language Learning Journal*, 31, 23–29.
<https://doi.org/10.1075/pbns.140.19mea>
- Milton, J. (2009). *Measuring second language vocabulary acquisition* (Vol. 45). Multilingual Matters.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge University Press. <https://doi.org/10.1017/9781009093873>

- Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review*, 63(1), 59–82.
<https://doi.org/10.3138/cmlr.63.1.59>
- Oscarson, A. D. (2009). *Self-assessment of writing in learning English as a foreign language: A study at the upper secondary school level* (Göteborg Studies in Educational Sciences No. 277). University of Gothenburg.
- Read, J. (2000). *Assessing vocabulary*. Cambridge University Press.
- Ross, S. (1998). Self-assessment in second language testing: A meta-analysis and analysis of experiential factors. *Language Testing*, 15(1), 1–20.
<https://doi.org/10.1177/026553229801500101>
- Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Palgrave Macmillan. <https://doi.org/10.1057/9780230293977>
- Webb, S. (2005). Receptive and productive vocabulary learning: The effects of reading and writing on word knowledge. *Studies in Second Language Acquisition*, 27(1), 33–52. <https://doi.org/10.1017/S0272263105050023>
- Webb, S., & Nation, I. S. P. (2017). *How vocabulary is learned*. Oxford University Press.
- Young, A., & Son, J. B. (2023). Synchronous computer-mediated communication and task-based learning in the EFL classroom. *Language Teaching Research*. Advance online publication. <https://doi.org/10.1177/13621688231191309>