Vocabulary Knowledge and Metacognitive Awareness in L2 Listening: Testing the Core-Peripheral Hypothesis

Dan Li[1], Apisak Sukying[1], & Nithipong Yothachai[2]

[1] Department of Western Languages and Linguistics, Faculty of Humanities and Social Sciences, Mahasarakham University, Maha Sarakham, 44150, Thailand

[2] Department of English Education, Faculty of Education and Human Development, Chaiyaphum Rajabhat University, Chaiyaphum, 36000, Thailand

Correspondence: Apisak Sukying, Department of Western Languages and Linguistics, Faculty of Humanities and Social Sciences, Mahasarakham University, Maha Sarakham, Thailand 44150. E-mail: apisak.s@msu.ac.th

Received: September 5, 2025 Accepted: November 3, 2025 Online Published: December 23, 2025

doi:10.5430/wjel.v16n3p38 URL: https://doi.org/10.5430/wjel.v16n3p38

Abstract

This study tested predictions from Hulstijn's core-peripheral framework for second-language listening by comparing the relative contributions of vocabulary knowledge and metacognitive awareness among Chinese EFL learners. Participants were 166 third-year English majors with intermediate proficiency (TEM-4 scores 65–75). They completed comprehensive assessments of four vocabulary dimensions: written breadth, written depth, aural breadth, and aural depth. They also completed a measure of metacognitive awareness and a standardized listening test. Hierarchical regression analysis revealed that vocabulary knowledge accounted for the majority of variance in listening. Aural vocabulary measures explained substantial additional variance beyond written measures. By contrast, metacognitive awareness contributed only minimal incremental variance after vocabulary had been entered into the model. Within the vocabulary dimensions, aural depth and aural breadth were the strongest predictors of performance. The coefficients for written measures were markedly reduced once aural measures were statistically controlled. These results support a hierarchical distinction between core and peripheral components. They position vocabulary knowledge, particularly in its aural modality, as foundational to listening success, whereas metacognitive awareness plays a supplementary role that may vary across proficiency levels. The clear superiority of aural over written vocabulary challenges assessment practices that prioritize orthographic knowledge, underscoring the need for modality-specific vocabulary instruction to enhance L2 listening comprehension.

Keywords: core-peripheral hypothesis, L2 listening comprehension, vocabulary knowledge, aural vocabulary, metacognitive awareness.

1. Introduction

The architecture of second language (L2) listening comprehension has long fascinated researchers who seek to distinguish fundamental from facilitative components. Vocabulary knowledge consistently emerges as a robust predictor of listening performance (Matthews, 2018; Matthews & Cheng, 2015; Stæhr, 2009; Wang & Treffers-Daller, 2017). By contrast, estimates of the impact of metacognitive awareness remain mixed, with effect sizes that vary considerably across studies (Vandergrift & Baker, 2015; Wallace, 2022). This pattern suggests a hierarchical organization of the listening construct in which some components operate as prerequisites and others provide enhancement.

Hulstijn's (2015) core-peripheral framework provides a principled account of this hierarchy by defining core components as automatized linguistic knowledge essential for basic comprehension and peripheral components as factors that facilitate performance without determining it. Core components, especially vocabulary and grammar, form the foundation of language processing across modalities. Peripheral components, including metacognitive strategies and pragmatic knowledge, may improve performance under favorable conditions but do not compensate for deficits in core knowledge. From this perspective, the framework predicts systematically stronger associations between vocabulary knowledge and listening comprehension than between metacognitive awareness and listening outcomes.

Despite the framework's appeal, relatively few empirical studies have examined vocabulary and metacognition together in L2 listening. Many investigations have treated these constructs separately (Masrai, 2022; Matthews, 2018; Matthews & Cheng, 2015), or have not distinguished between written and aural vocabulary modalities (Vandergrift & Baker, 2015; Wang & Treffers-Daller, 2017). This lack of differentiation is problematic because psycholinguistic evidence indicates that aural and written vocabulary constitute distinct dimensions of knowledge with different relationships to listening proficiency (Cheng & Matthews, 2018; Ha, 2021).

The present study addresses these gaps by examining four dimensions of vocabulary knowledge, written breadth, written depth, aural breadth, and aural depth, together with metacognitive awareness as predictors of L2 listening comprehension. We test whether the core–peripheral distinction holds across modalities and whether aural measures show superior predictive validity, as processing-based accounts suggest (Field, 2019). This design permits a rigorous test of the hierarchical relations proposed in Hulstijn's framework and clarifies the relative contributions of modality-specific vocabulary knowledge to listening success.

2. Literature Review

2.1 Vocabulary Knowledge as a Core Component

Vocabulary knowledge has long been recognized as foundational to second language (L2) listening comprehension. Stæhr (2009) demonstrated that vocabulary size accounted for approximately 50% of the variance in listening performance among Danish EFL learners, thereby establishing the central role of vocabulary in this context. Subsequent research refined this understanding by distinguishing between vocabulary breadth (the number of words known) and depth (the quality of word knowledge, including collocations, polysemy, and semantic associations). A recent meta-analysis by Zhang and Zhang (2022), which synthesized 88 empirical studies, confirmed vocabulary knowledge as a robust predictor of both L2 listening and reading comprehension. The mean correlations reached r = .56 for listening and r = .63 for reading, underscoring the foundational role of vocabulary across receptive skills.

The breadth–depth distinction is particularly important in listening contexts. Wang and Treffers-Daller (2017) found that vocabulary depth, measured with word association tasks, uniquely predicted inferential listening comprehension beyond breadth among Chinese learners. Deep lexical knowledge, in particular knowledge of collocations and semantic relations, enables the processing of implicit meanings and the generation of inferences during real-time listening. Han and Qian (2024) extended these findings by showing that breadth provides a necessary foundation at lower proficiency levels. In contrast, depth becomes more important for advanced learners who are processing complex discourse. Evidence from EFL writing contexts further demonstrates the pronounced role of vocabulary depth in productive language use. Sukying (2023) reported that vocabulary depth, measured with the Productive Vocabulary Levels Test (PVLT) and the Word Associates Test (WAT), explained 71.2% of the variance in L2 argumentative writing performance among Thai postgraduate students and substantially outweighed vocabulary breadth. This pattern suggests that breadth provides essential lexical resources, whereas depth is particularly crucial for advanced productive use, which requires lexical precision, semantic sophistication, and appropriate collocational knowledge.

Beyond the breadth–depth dimension, accumulating evidence indicates that aural and written vocabulary constitute partially distinct knowledge stores. Matthews and Cheng (2015) showed that recognition of high-frequency words from speech predicted listening comprehension more strongly than written vocabulary measures, with aural vocabulary correlations of r = .67–.73. Masrai (2019, 2020) provided converging evidence that phonological word knowledge predicts listening performance more robustly than orthographic knowledge, suggesting that modality-matched assessment captures the lexical resources most relevant to listening comprehension. Cheng and Matthews (2018) subsequently confirmed that three vocabulary measures, written recognition, aural recognition, and meaning recall, exhibited different relationships with listening, with aural measures showing superior predictive power. Experimental evidence further supports this modality-specific effect: Uchihara (2023) found that aural vocabulary testing produced superior learning of spoken word forms compared with written testing, which suggests that assessment modality directly shapes the development of modality-specific lexical representations.

The skill specificity of vocabulary contributions extends beyond the aural–written distinction. Sukying (2025) demonstrated that different vocabulary measures differentially predicted speaking and writing abilities among Thai EFL undergraduates: form–meaning knowledge (VLT and WAT) explained 28.8% of the variance in speaking, whereas productive depth measures (PVLT and WAT) explained 71.2% of the variance in writing. This pattern suggests that vocabulary assessment should align with the processing demands of the target skills. This insight is particularly relevant for listening research in which rapid phonological access differs fundamentally from the controlled lexical retrieval processes that characterize writing.

These modality effects reflect fundamental differences in lexical processing. During listening, learners must rapidly activate phonological representations under severe time constraints and without opportunities for reinspection (Field, 2019). Written vocabulary knowledge, which is acquired predominantly through orthographic input in many EFL contexts, may not automatically transfer to listening situations that require direct phonological access. The dissociation between orthographic and phonological lexicons is therefore a critical consideration for listening comprehension research and pedagogy. The divergent vocabulary demands across skills, with written depth dominating in writing (Sukying, 2023), with form-meaning links contributing to speaking (Sukying, 2025), and with aural measures predicting listening (Matthews & Cheng, 2015), underscore the multidimensional nature of vocabulary knowledge and highlight the need for modality-matched assessment approaches.

2.2 Metacognitive Awareness as a Peripheral Component

Metacognitive awareness in L2 listening encompasses learners' declarative knowledge of listening processes, along with their procedural regulation through planning, monitoring, and evaluation (Vandergrift & Goh, 2012). The Metacognitive Awareness Listening Questionnaire (MALQ; Vandergrift et al., 2006) operationalises this construct across five dimensions: problem-solving, planning–evaluation, directed attention, person knowledge (self-efficacy beliefs), and mental translation, which is reverse-scored because it is considered counterproductive.

Although metacognitive awareness correlates positively with listening performance, its predictive power is typically substantially weaker than that of vocabulary knowledge. Vandergrift and Baker (2015) argued that metacognitive awareness influences listening primarily indirectly by facilitating vocabulary acquisition rather than by exerting a direct effect on comprehension. Consistent with this view, Wallace (2022) reported that metacognitive awareness explained only modest additional variance beyond vocabulary knowledge, a pattern that aligns with peripheral rather than core status.

Cognitive capacity constraints offer one explanation for the limited contribution of metacognitive awareness. When core linguistic processing remains effortful and resource-intensive, fewer attentional resources are available for metacognitive monitoring and strategic deployment (Anderson, 2015). Recent studies provide converging evidence for this proficiency-dependent pattern. Milliner and Dimoski (2024) found that lower-proficiency learners (CEFR A2) who received process-based metacognitive instruction showed only modest gains in listening self-efficacy despite improvements in listening proficiency, which suggests a restricted capacity to implement metacognitive strategies when linguistic resources are constrained. Muhammadpour et al. (2024) further showed that working-memory capacity, a proficiency-related cognitive resource, moderates the effectiveness of metacognitive interventions, with high-capacity learners achieving substantially larger gains than low-capacity learners.

Diagnostic research further clarifies how proficiency influences the effectiveness of metacognitive strategies. Using cognitive diagnostic modeling with 567 Chinese EFL learners, Fu et al. (2023) reported that metacognitive factors correlate differently with higher-order and lower-order listening subskills. Mental translation showed negative correlations across all subskills, with values ranging from $r = -.165$ to $-.035$, indicating a counterproductive strategy used more frequently by lower-proficiency learners. This pattern suggests that less skilled listeners rely more on ineffective strategies, whereas more skilled listeners deploy metacognitive strategies more automatically and efficiently. This interpretation reconciles apparently contradictory findings: at lower proficiency levels, where core knowledge is limited, metacognitive strategies cannot compensate for insufficient linguistic resources; at intermediate and advanced levels, where core processing is more automatized, metacognitive strategies yield more substantial benefits (Du & Man, 2023). Overall, the developmental trajectory suggests that metacognitive awareness serves as a facilitative enhancement, amplifying existing linguistic competence rather than as a compensatory mechanism that can substitute for deficits in core knowledge.

2.3 The Core–Peripheral Framework Applied to Listening

Hulstijn’s (2015) core–peripheral framework distinguishes core components (phonological, morphosyntactic, and lexical knowledge) from peripheral components (strategic competence and metalinguistic knowledge). He argued that peripheral components cannot operate independently of core components, whereas core components can function independently. Applied to L2 listening comprehension, this framework predicts that vocabulary knowledge should show stronger predictive relationships with listening proficiency than metacognitive awareness. Metacognitive strategies are therefore expected to contribute only modest incremental variance after lexical resources are controlled (Wallace, 2022; Wang & Treffers-Daller, 2017).

The framework’s hierarchical prediction aligns with processing models that emphasize bottom-up lexical access as a prerequisite for higher-level strategic application. Evidence from multiple sources supports this theoretical positioning. Wang and Treffers-Daller (2017) found that vocabulary knowledge and general language proficiency each explained 13% of the variance in listening comprehension among Chinese learners studying at UK universities. This proportion was substantially greater than that associated with metacognitive awareness, which contributed a smaller amount of additional variance. They further argued that metacognitive awareness functions as a “domain-general peripheral factor” that operates indirectly through core knowledge rather than as a primary driver of listening success. Wallace (2022) extended these findings by showing that metacognitive awareness explained only modest additional variance beyond vocabulary knowledge, with effect sizes consistent with peripheral rather than core status.

Recent research further refines the framework by suggesting that the core–peripheral distinction may manifest differently across proficiency levels. Zhang and Graham (2020) demonstrated that listening proficiency outweighed pre-existing vocabulary knowledge in vocabulary learning through aural input, indicating complex bidirectional relationships between core components and listening outcomes. In addition, the proficiency-dependent patterns reviewed in Section 2.2 indicate a skill-level asymmetry: metacognitive awareness has limited utility for learners with lower proficiency. Still, it yields greater benefits for intermediate and advanced learners. Taken together, these observations suggest that the contribution of peripheral components increases as core components become more automatized. This developmental trajectory is consistent with skill-acquisition theory: as core linguistic processing shifts from controlled to automatic, cognitive resources become available for the strategic deployment of peripheral processes.

The framework also requires extension to accommodate modality-specific realizations of core components. Although Hulstijn (2015) treats lexical knowledge as core without distinguishing input modalities, a growing body of evidence indicates that phonological and orthographic vocabulary constitute partially distinct knowledge stores (Cheng & Matthews, 2018; Matthews & Cheng, 2015). We propose that phonological vocabulary knowledge serves as the core for listening and speaking, whereas orthographic vocabulary knowledge serves as the core for reading and writing. This refinement preserves the framework’s hierarchical insight within the core–peripheral perspective by separating core linguistic knowledge from peripheral strategic competence and by recognizing internal differentiation within core components that aligns with processing modality. The relatively weak correlations across modalities reported in prior research support this modality-specific conceptualization.

However, empirical tests that incorporate both modality-specific vocabulary measures and metacognitive awareness remain rare, which limits our ability to evaluate the framework’s predictions comprehensively across proficiency levels and input modalities. The present study addresses this gap by examining vocabulary dimensions across modalities together with metacognitive awareness, thereby providing a rigorous empirical test of the core–peripheral distinction in L2 listening. By sampling learners at intermediate proficiency (TEM-4 scores 65–75), the design enables us to examine the framework’s predictions at a developmental stage where both core and peripheral components are likely to be measurably active. This approach avoids floor effects that can obscure relationships at lower proficiency levels and reduces the risk of ceiling effects at advanced levels.

3. Method

3.1 Research Design

This study employed a quantitative cross-sectional design to investigate the concurrent relationships among vocabulary knowledge dimensions, metacognitive awareness, and L2 listening comprehension. Guided by multidimensional perspectives on vocabulary acquisition (Nation, 2013; Schmitt, 2014; Schmitt & Schmitt, 2014) and by Field's (2019) modality-specific processing account, we administered a comprehensive battery of instruments to test whether modality-matched vocabulary measures exhibit stronger associations with listening outcomes than traditional written assessments. This design enabled an empirical test of predictions derived from Hulstijn's (2015) core–peripheral framework while also addressing methodological limitations in prior research.

3.2 Participants

Participants were 166 third-year English majors (128 female, 38 male; age range, 19–21 years; **M** = 20.3, **SD** = 0.8), recruited from a regional public university in eastern China. All had completed approximately 10–13 years of prior English instruction and four semesters of listening-focused coursework (Intensive Listening, Extensive Listening, and Integrated English Skills). Participants' scores on the Band 4 level of the Test for English Majors (TEM-4) ranged from 65 to 75 (**M** = 69.2, **SD** = 3.4), corresponding to intermediate proficiency on nationally recognized benchmarks (Jin & Fan, 2011; Qian & Lin, 2020). One participant with recent IELTS preparation experience was excluded to minimize practice effects.

A priori power analysis using G*Power 3.1 (Faul et al., 2009) indicated that the sample size was adequate to detect medium effects with five predictors, assuming ($\mathbf{1-\beta}$) = .80, $\mathbf{f}^2$= .15, and $\boldsymbol{\alpha}$ = .05. The study received institutional ethics approval, and all participants provided written informed consent.

3.3 Instruments

3.3.1 Written Vocabulary Breadth

The Updated Vocabulary Levels Test (VLT; Webb et al., 2017) assessed receptive vocabulary size across five frequency bands (1,000-5,000 word families). The test employed a matching format with 30 items at each frequency level (maximum score: 150). It incorporated contemporary corpus data from the British National Corpus and the Corpus of Contemporary American English to ensure current relevance. Internal consistency reliability was strong ($\boldsymbol{\alpha}$ = .85).

3.3.2 Written Vocabulary Depth

Read's (1998) Word Associates Test (WAT) measured the quality of lexical knowledge with 40 items that targeted both paradigmatic (synonym) and syntagmatic (collocational) relations. Each item presented a target word with eight response options, and participants selected four semantically or collocationally related associates. Target words were sampled from the 2,000–3,000 frequency bands, yielding a maximum score of 160 (one point per correct associate). Validation research has demonstrated the test's reliability and construct validity for measuring depth of vocabulary knowledge (Qian, 1999, 2002). In the present sample, internal-consistency reliability was good ($\boldsymbol{\alpha}$ = .87).

3.3.3 Aural Vocabulary Breadth

A-LEX (Milton & Hopkins, 2005) assessed phonological vocabulary recognition with 120 aurally presented real words distributed across five frequency bands (1,000–5,000) and 20 pseudowords used to detect and adjust for response bias. Participants listened to each item once via headphones and made binary yes/no recognition judgments within a fixed 3-second response window. All audio stimuli were produced by a native speaker of British English. Scoring incorporated proportional penalties for pseudoword false alarms, following the protocol of Milton and Hopkins (2005). Internal-consistency reliability was acceptable ($\boldsymbol{\alpha}$ = .86), which is comparable to coefficients reported in validation studies (range $\boldsymbol{\alpha}$ = .82–.89) across diverse learner populations (Masrai, 2020, 2022; Milton & Hopkins, 2006).

3.3.4 Aural Vocabulary Depth

An adapted aural version of the Word Associates Test (A-WAT) mirrored the written instrument's structure by presenting 40 items that assessed both paradigmatic (synonym) and syntagmatic (collocational) knowledge through phonological input. For each item, participants heard a target word twice for clarity, followed by eight options presented auditorily in randomized order; each option was repeated once. From these options, participants selected four correct associates. Target words matched the written version's difficulty range (2,000–3,000 frequency bands), which enabled direct cross-modal comparison. All stimuli were recorded by a native speaker of Canadian English with doctoral training in linguistics to ensure phonological accuracy. Participants responded at their own pace to minimize processing demands unrelated to lexical knowledge. The adapted instrument showed strong internal-consistency reliability ($\boldsymbol{\alpha}$ = .85), which is comparable to reliability coefficients reported for written association-based depth measures (Qian, 1999, 2002; Read, 1998).

3.3.5 Metacognitive Awareness

The Metacognitive Awareness Listening Questionnaire (MALQ; Vandergrift et al., 2006) measured learners' strategic knowledge and self-regulation during listening comprehension. The expanded 30-item version retained the original five-factor structure, with the following subscales: problem solving, planning and evaluation, directed attention, person knowledge, and mental translation (reverse-scored). Participants responded on a six-point Likert scale, ranging from 1 (strongly disagree) to 6 (strongly agree), with higher scores indicating greater metacognitive awareness.

We administered a bilingual Chinese–English version that was developed with established translation procedures. A professional translator conducted the forward translation, an independent bilingual researcher completed the back translation, and two native Mandarin-speaking instructors provided the final review for pragmatic clarity and cultural appropriateness. Reliability coefficients ranged from acceptable to good: overall scale (α = .83), with subscales from α = .65 for Mental Translation to α = .73 for Problem Solving, consistent with psychometric properties reported in international validation studies (Vandergrift et al., 2006; Vandergrift & Goh, 2012).

3.3.6 Listening Comprehension

An official IELTS Academic Listening form (Taylor & Falvey, 2007) provided the criterion measure of L2 listening proficiency. This 40-item standardised assessment sampled systematically across discourse types (social dialogue, monologic presentation, academic discussion, and formal lecture) and task formats (multiple-choice, matching, gap-fill, sentence completion, and diagram labelling). The 30-minute audio presentation was followed by a 10-minute transfer period for recording final answers.

Each correct response earned one point, and raw scores (ranging from 0 to 40) were retained for all analyses to maximise measurement precision. Test administration followed standardised IELTS protocols, including strict timing, audio quality verification, and invigilated conditions to ensure score validity. The assessment yielded acceptable internal-consistency reliability (α = .79), which is appropriate for a heterogeneous criterion measure that samples multiple listening subcompetencies.

3.4 Procedure

Data collection took place over two consecutive days during regular class times. On Day 1, participants completed the written vocabulary assessments (VLT and WAT), took a 20-minute break, and then proceeded to the aural vocabulary tests (A-LEX and A-WAT). On Day 2, they first completed the IELTS Listening test and then completed the MALQ. This sequencing ensured that metacognitive self-reports were obtained immediately after the substantive listening task performance. All assessments were administered in groups with standardized instructions under invigilated conditions.

3.5 Data Analysis

All analyses were conducted in IBM SPSS Statistics, version 27, with a significance level set at α = .05 (two-tailed). Descriptive statistics and distributional properties were examined using conventional thresholds, |skewness| < 2 and |kurtosis| < 7 (Hair et al., 2019). We used Pearson correlations to examine bivariate associations among all variables.

Hierarchical multiple regression was used to quantify the unique and incremental contributions of vocabulary dimensions and metacognitive awareness to listening comprehension. This approach allows theory-driven ordering of predictors and yields explicit estimates of incremental variance at each step (Keith, 2015); therefore, it was preferred to simultaneous or stepwise entry. For Research Question 1, two complementary models were tested: a modality-based sequence (in which written vocabulary is entered before aural vocabulary) and a dimension-based sequence (in which breadth is entered before depth). For Research Question 2, all vocabulary measures were entered first to establish a lexical baseline, and the MALQ total score was then added in the second step to estimate the incremental contribution of metacognitive awareness.

Effect sizes were expressed as Cohen's f^2 with benchmarks of .02, .15, and .35 for small, medium, and large effects, respectively (Cohen, 1988). Incremental effects for added blocks were computed as $f^2 = \Delta R^2/(1 - R^2_{full})$. Multicollinearity was assessed using tolerance and VIF with acceptable thresholds of tolerance > .10 and VIF < 10 (Hair et al., 2019). Parametric assumptions were evaluated in accordance with established guidelines (Tabachnick & Fidell, 2019): linearity, homoscedasticity, independence of errors, and normality of residuals. No substantial violations were detected.

4. Results

4.1 Vocabulary Dimensions as Predictors of L2 Listening Comprehension

4.1.1 Descriptive Statistics and Preliminary Analyses

Comprehensive descriptive analyses examined distributional properties for all study variables prior to hypothesis testing. Table 1 presents descriptive statistics for vocabulary measures, metacognitive awareness components, and listening proficiency.

Table 1. Descriptive Statistics for All Study Variables (N = 166)

Variable	Possible Range	Observed Range	Mean (%)	SD	Skewness	Kurtosis
Vocabulary Knowledge Measures						
VLT (Written Breadth)	0–150	83–134	103.87 (69.25%)	8.05	0.09	0.89
WAT (Written Depth)	0–160	96–138	110.21 (68.88%)	10.43	0.68	–0.09
A-LEX (Aural Breadth)	0–120	69–106	87.82 (73.18%)	7.39	0.13	–0.12
A-WAT (Aural Depth)	0–160	67–113	89.81 (56.13%)	9.05	0.18	–0.28
Metacognitive Awareness (MALQ)						
Problem-Solving	9–54	29–54	38.84 (71.93%)	4.67	0.15	0.11
Directed Attention	5–30	10–29	17.30 (57.67%)	4.56	0.21	–0.57
Planning-Evaluation	6–36	10–30	17.63 (48.97%)	4.83	0.38	–0.84
Person Knowledge	6–36	10–35	21.26 (59.06%)	6.25	0.32	–0.53
Mental Translation	4–24	4–20	10.80 (45.00%)	2.89	0.57	0.98
Total Metacognitive Awareness	30–180	75–137	105.83 (58.79%)	14.31	0.36	–0.63
Listening Proficiency						
IELTS Listening	0–40	21–35	28.22 (70.55%)	2.78	0.10	–0.19

Note. VLT = Vocabulary Levels Test (written breadth); WAT = Word Associates Test (written depth); A-LEX = Aural Lex (aural breadth); A-WAT = Aural Word Associates Test (aural depth); MALQ = Metacognitive Awareness Listening Questionnaire. Percentages are calculated as (Mean/Maximum possible score) × 100. Skewness and kurtosis values fall within commonly used ranges (±2 and ±7, respectively), supporting the use of parametric analyses.

Participants' vocabulary knowledge showed adequate variability across all measures, with mean performance ranging from 56.13% on the A-WAT to 73.18% on the A-LEX, calculated as a percentage of the maximum possible score. Listening proficiency averaged 28.22 out of 40 (70.55% accuracy), which is consistent with the intermediate-to-advanced proficiency typically observed among third-year English majors in Chinese universities (Zhang & Graham, 2020). All variables met the distributional criteria for parametric analyses, with skewness values ranging from 0.09 to 0.68 and kurtosis values ranging from −0.84 to 0.98, both within commonly accepted ranges (±2 and ±7, respectively; Hair et al., 2019).

4.1.2 Correlation Analyses

Pearson product-moment correlations were used to examine bivariate relationships between vocabulary dimensions and listening comprehension. Table 2 presents the complete correlation matrix.

Table 2. Correlations among vocabulary measures and listening proficiency (N = 166)

	VLT	WAT	A-LEX	A-WAT	IELTS
VLT	--				
WAT	.609**	--			
A-LEX	.299**	.408**	--		
A-WAT	.269**	.376**	.711**	--	
IELTS	.456**	.502**	.624**	.617**	--

Note. VLT = Vocabulary Levels Test (written receptive vocabulary breadth); WAT = Word Associates Test (written receptive vocabulary depth); A-LEX = Aural Vocabulary Levels Test (aural receptive vocabulary breadth); A-WAT = Aural Word Associates Test (aural receptive vocabulary depth); IELTS = listening proficiency. **p < .01 (two-tailed).

The correlation analyses revealed modality-specific patterns. Aural vocabulary measures displayed substantially stronger associations with listening proficiency than written measures. A-LEX showed the strongest correlation with IELTS Listening (**r** = .624, **p** < .001; $\mathbf{r}^2$= .389), followed closely by A-WAT (**r** = .617, **p** < .001; $\mathbf{r}^2$= .381). Taken together, the aural measures explained approximately 39% of the variance in listening, which was about 68% higher than the average for the written measures (≈23%).

Written measures, although statistically significant, showed weaker relationships with listening: WAT correlated moderately with listening (**r** = .502, **p** < .001; $\mathbf{r}^2$= .252), and VLT had the weakest association (**r** = .456, **p** < .001; $\mathbf{r}^2$= .208). This pattern highlights the crucial role of phonological lexical knowledge in listening comprehension.

Intercorrelations were moderate within the written measures (VLT–WAT: **r** = .609, **p** < .001) and strong within the aural measures (A-LEX–A-WAT: **r** = .711, **p** < .001), whereas cross-modal correlations were weaker, ranging from **r** = .269 to **r** = .408. These patterns suggest that written and aural vocabulary constitute related yet distinct dimensions, consistent with accounts of modality-specific lexical representation (Cheng & Matthews, 2018).

4.1.3 Hierarchical Regression Analyses

Hierarchical multiple regression examined the unique and combined predictive power while controlling for shared variance. Variables entered in theoretically meaningful blocks: written vocabulary first to establish baseline contributions, followed by aural vocabulary to test incremental validity. Table 3 presents the regression results.

Table 3. Hierarchical Regression Analysis: Vocabulary Measures Predicting Listening Proficiency (N = 166)

Model & Predictor	B	SE B	β	P	R^2	ΔR^2	f ²(model)	Δf^2
Model 1: Written Vocabulary								
(Constant)	9.211	2.458	—	.000	.288	.288	0.405	0.405
VLT	.082	.029	.238	.005				
WAT	.095	.022	.357	.000				
Model 2: Written + Aural Vocabulary								
(Constant)	-1.099	2.325	—	.637	.531	.244	1.132	0.520
VLT	.068	.024	.196	.005				
WAT	.040	.019	.151	.036				
A-LEX	.109	.030	.289	.000				
A-WAT	.092	.024	.301	.000				

Note. VLT = Vocabulary Levels Test (written breadth); WAT = Word Associates Test (written depth); A-LEX = aural lexical recognition test (aural breadth); A-WAT = Aural Word Associates Test (aural depth). Unstandardized coefficients (B) are reported with standard errors (SE B); standardized coefficients are denoted by β. Two-tailed p values are shown. Model effect sizes were computed as $f^2 = R^2/(1 - R^2)$; incremental effect sizes were computed as $\Delta f^2 = \Delta R^2/(1 - R^2$, full model). Benchmarks for f^2: small = .02, medium = .15, large = .35 (Cohen, 1988). N = 166.

Model 1, which contained the written vocabulary measures (VLT and WAT), was statistically significant, **F**(2, 163) = 32.910, **p** < .001, and accounted for 28.8% of the variance in listening comprehension (adjusted R^2= .279), representing a large effect (f^2= 0.405). Within this model, WAT was the stronger predictor (**β** = .357, **p** < .001) compared with VLT (**β** = .238, **p** = .005).

Model 2 added the aural vocabulary measures, yielding R^2= .531 (adjusted R^2= .520) and an additional 24.4% of explained variance; the change test was significant, **F**(2, 161) = 41.873, **p** < .001. The incremental effect size for the added block was f^2= 0.520, indicating that phonological vocabulary knowledge explains substantial variance beyond orthographic knowledge. The overall model effect size reached f^2 = 1.132, which indicates very large practical significance.

In the full model, A-WAT was the strongest single predictor (**β** = .301, **p** < .001), followed by A-LEX (**β** = .289, **p** < .001). Notably, the standardized coefficients for the written measures decreased, VLT to **β** = .196 and WAT to **β** = .151, indicating that much of their initial predictive power was subsumed by the aural measures. This pattern suggests that aural vocabulary captures unique variance in listening comprehension that written measures do not capture.

To verify that the modality effect was not dependent on entry order, an alternative hierarchical regression entered breadth measures (VLT and A-LEX) first, followed by depth measures (WAT and A-WAT). Table 4 presents these results.

Table 4. Hierarchical Regression Analysis: Alternative Entry Order (Breadth then Depth) (N = 166)

Model & Predictor	B	SE B	β	*P*	R^2	ΔR^2	f ²(model)	Δf^2
Model 1: Breadth Measures (Written + Aural)								
(Constant)	-.056	2.448	—	.982	.469	.469	0.883	0.883
VLT	.102	.021	.296	.000				
A-LEX	.201	.022	.535	.000				
Model 2: Breadth + Depth Measures								
VLT	.068	.024	.196	.005	.531	.063	1.132	0.134
A-LEX	.109	.030	.289	.000				
WAT	.040	.019	.151	.036				
A-WAT	.092	.024	.301	.000				

Note. *p* < .001 (two-tailed).

When breadth measures were entered first (Model 1), they explained 46.9% of the variance in listening comprehension (adjusted R^2= .462), **F**(2, 163) = 71.914, **p** < .001, with A-LEX (**β** = .535) substantially outweighing VLT (**β** = .296). Model 2 added the depth measures, increasing R^2 to .531 (ΔR^2= .063); the change test was significant, **F**change(2, 161) = 10.761, **p** < .001. Although statistically significant, this modest increment (Δf^2= 0.134) suggests that, once breadth is controlled, depth provides supplementary rather than transformative predictive power.

A critical examination of standardised coefficients in the full model reveals that modality distinctions take precedence over breadth–depth distinctions. Both aural measures, breadth (**β** = .289) and depth (**β** = .301), consistently outperformed their written counterparts, namely breadth (**β** = .196) and depth (**β** = .151). This pattern suggests that phonological knowledge, whether measured in terms of breadth or depth, is more consequential for listening comprehension than the specific dimension of lexical knowledge.

4.1.4 Multicollinearity Diagnostics

Given moderate intercorrelations among predictors, multicollinearity diagnostics proved essential. Table 5 presents tolerance and VIF values.

Table 5. Collinearity diagnostics for vocabulary measures (N= 166)

Vocabulary Measures	Tolerance	Variance inflation factor (VIF)
VLT	.626	1.598
A-LEX	.470	2.128
WAT	.567	1.765
A-WAT	.485	2.061

All tolerance values exceeded 0.470, which is well above the critical threshold of 0.10, and all VIF values were below 2.128, which is far below the conventional cutoff of 10. These diagnostics indicate that the predictors are sufficiently independent, supporting the validity and stability of the regression coefficients.

4.1.5 Summary for Research Question One

Vocabulary dimensions collectively explained 53.1% of the variance in L2 listening comprehension, supporting the status of vocabulary knowledge as a core component. A clear predictive hierarchy emerged: aural measures contributed an additional 24.4% of explained variance beyond written measures ($\Delta \mathbf{R}^2 = .244$), whereas written measures contributed only 6.3% beyond aural measures ($\Delta \mathbf{R}^2 = .063$). This asymmetry highlights the primacy of phonological lexical knowledge. Relatively weak cross-modal correlations, in contrast to strong within-modality correlations, indicate that the two modalities represent related but distinct knowledge stores that develop at least partly independently, challenging assumptions about automatic transfer between modalities.

4.2 Metacognitive Awareness and L2 Listening Comprehension

4.2.1 Correlation Analyses

Pearson correlations were used to examine the relationships between metacognitive components and listening proficiency. Table 6 presents the correlation matrix.

Table 6. Correlations among metacognitive components and listening proficiency (N = 166)

	IELTS	PS	DA	PE	PK	MT
IELTS	--					
PS	.430**	--				
DA	.400**	.444**	--			
PE	.407**	.450**	.433**	--		
PK	.385**	.428**	.349**	.389**	--	
MT	-.215**	-.247**	-.211**	-.313**	-.395**	--

Note. IELTS = listening proficiency; PS = problem solving; DA = directed attention; PE = person knowledge; PK = planning knowledge; MT = mental translation. **p < .01 (two-tailed).

Problem Solving showed the strongest relationship with listening proficiency (**r** = .430, **p** < .001; $\mathbf{r}^2 = .185$), followed by Planning and Evaluation (**r** = .407, **p** < .001; $\mathbf{r}^2 = .166$) and Directed Attention (**r** = .400, **p** < .001; $\mathbf{r}^2 = .160$). Person Knowledge showed a somewhat weaker association (**r** = .385, **p** < .001; $\mathbf{r}^2 = .148$). Mental Translation was negatively correlated with listening (**r** = −.215, **p** = .005), indicating that reliance on L1 translation during listening may impair comprehension. Such reliance can create processing bottlenecks in which learners miss subsequent input while translating prior segments (Vandergrift et al., 2006).

Intercorrelations among the adaptive strategies (Problem Solving, Planning and Evaluation, and Directed Attention) ranged from **r** = .433 to .450, suggesting that these strategies function as an integrated strategic repertoire. Mental Translation correlated negatively with all other components, ranging from **r** = −.211 to −.395, which confirms that it operates as a counterproductive strategy distinct from the adaptive metacognitive processes.

4.2.2 Hierarchical Regression Analysis

Hierarchical regression was used to test whether metacognitive awareness adds predictive power beyond vocabulary knowledge. Table 7 presents sequential models.

Table 7. Hierarchical Regression Analysis: Vocabulary and Metacognitive Awareness Predicting Listening Proficiency (N = 166)

Model & Predictor	B	SE B	β	P	R ²	ΔR²	f ²(model)	Δf²
Model 1: Vocabulary Knowledge Only								
(Constant)	-1.099	2.325	—	.637	.531	.531	1.132	1.132
VLT	.068	.024	.196	.005				
WAT	.040	.019	.151	.036				
A-LEX	.109	.030	.289	.000				
A-WAT	.092	.024	.301	.000				
Model 2: Vocabulary + Total Metacognitive Awareness								
(Constant)	-.195	2.299	—	.933	.553	.022	1.237	0.049
VLT	.057	.023	.165	.016				
WAT	.014	.021	.053	.496				
A-LEX	.102	.029	.271	.001				
A-WAT	.088	.023	.285	.000				

Total_Meta	.039	.014	.201	.006

Note. Model 1 includes all four vocabulary measures; Model 2 adds the total metacognitive awareness score. Total_Meta = composite score from MALQ (Problem-Solving, Directed Attention, Planning-Evaluation, Person Knowledge, Mental Translation). Same abbreviations and effect size interpretations as Table 3. **p < .01, ***p < .001 (two-tailed).

Model 1 replicated the vocabulary analysis and explained 53.1% of the variance (adjusted $\mathbf{R}^2 = .520$), $\mathbf{F}(4, 161) = 45.644$, $\mathbf{p} < .001$, with the aural measures again emerging as the strongest predictors (A-WAT: $\boldsymbol{\beta} = .301$, $\mathbf{p} < .001$; A-LEX: $\boldsymbol{\beta} = .289$, $\mathbf{p} < .001$).

Model 2 added the total metacognitive-awareness score, yielding $\mathbf{R}^2 = .553$ (adjusted $\mathbf{R}^2 = .539$) and a 2.2% increase in explained variance ($\Delta\mathbf{R}^2 = .022$, $\mathbf{p} = .006$). Although statistically significant, this increment was small ($\Delta\mathbf{f}^2 = 0.049$), which supports the peripheral role of metacognitive awareness. The standardized coefficient for metacognition ($\boldsymbol{\beta} = .201$, $\mathbf{p} = .006$) indicates a real contribution that is nevertheless substantially smaller than the contribution of vocabulary (overall $\mathbf{f}^2 = 1.132$, large effect).

4.2.3 Multicollinearity Diagnostics

Table 8 presents collinearity diagnostics for metacognitive components.

Table 8. Collinearity diagnostics for metacognitive components

Metacognitive components	Tolerance	Variance inflation factor (VIF)
PS	.672	1.488
DA	.721	1.386
PE	.684	1.462
PK	.693	1.444
MT	.813	1.230
Total_Meta	.542	1.844

All tolerance values exceeded 0.542, which is well above the 0.10 threshold, and all VIF values were below 1.844, which is far below the conventional cutoff of 10. These diagnostics indicate no problematic multicollinearity and support the stability of the regression coefficients. They also justify treating the adaptive strategies and mental translation as distinct constructs.

4.2.4 Summary for Research Question Two

Metacognitive awareness correlated significantly with listening proficiency ($\mathbf{r} = .430$) yet added only 2.2% of explained variance beyond vocabulary knowledge ($\Delta\mathbf{R}^2 = .022$). This small effect ($\Delta^* f^{*2} = 0.049$) compared with vocabulary's large overall effect ($\mathbf{f}^2 = 1.132$) provides quantitative support for the **core–peripheral** distinction. Problem Solving, Planning and Evaluation, and Directed Attention function as an integrated set of adaptive strategies, whereas Mental Translation operates as a counterproductive process. Overall, metacognitive factors appear to be beneficial supplements rather than foundational determinants of listening comprehension.

5. Discussion

This study provides empirical support for Hulstijn's (2015) core-peripheral hypothesis in L2 listening. Vocabulary knowledge emerged as the dominant predictor, explaining 53.1% of the variance in listening comprehension, whereas metacognitive awareness contributed an additional 2.2%. These results indicate that vocabulary, particularly in its aural modality, is foundational to listening success, and that metacognition serves as a beneficial but peripheral support. Beyond confirming the hierarchical distinction, the findings also reveal modality-specific patterns that extend and refine the theoretical framework.

5.1 The Primacy of Aural Vocabulary Knowledge

The most striking finding is the clear superiority of aural over written vocabulary measures. Aural vocabulary contributed an additional 24.4% of explained variance beyond written measures, whereas written measures contributed only 6.3% beyond aural vocabulary. This fourfold asymmetry cannot be attributed solely to method variance, as both modalities used parallel instruments that assessed the same constructs (breadth and depth).

Cross-modal correlations were relatively weak, ranging from .269 to .408. In contrast, within-modality correlations were strong (greater than .60). These results suggest that written and aural vocabulary constitute partially distinct knowledge stores rather than a unified construct that transfers automatically across modalities. This pattern challenges Nation's (2022) view that vocabulary knowledge develops uniformly across input modes. Instead, our findings suggest that learners develop quasi-independent phonological and orthographic lexicons, with limited bidirectional transfer, at least at intermediate proficiency levels where instruction has emphasized written forms.

This modality dissociation aligns with contemporary psycholinguistic evidence on lexical representation. Cheng and Matthews (2018) showed that phonological vocabulary knowledge predicts listening through direct lexical-access routes, whereas orthographic knowledge requires additional phonological recoding steps. Field (2019) argues that successful listening depends on rapid activation of phonological representations under severe temporal constraints, in contrast to reading, where visual word forms remain available for reinspection and conscious reflection. During real-time aural input, learners must access meaning directly from ephemeral phonological signals without orthographic support, so written vocabulary knowledge offers only marginal benefit.

The pronounced advantage of aural vocabulary for predicting listening comprehension differs from patterns reported for other receptive and productive skills. Sukying (2023) demonstrated that written productive depth measures (PVLT and WAT) accounted for 71.2% of the variance in L2 argumentative writing among Thai postgraduate students, substantially exceeding the contribution of breadth. Similarly, Sukying (2025) found that form–meaning knowledge (VLT and WAT) explained 28.8% of the variance in speaking, whereas written productive depth measures (PVLT and WAT) explained 71.2% of the variance in writing among Thai EFL undergraduates. This skill-specific pattern, with aural vocabulary dominating listening and written vocabulary depth predicting writing, supports the claim that lexical processing demands are modality matched.

The divergent vocabulary requirements across language skills likely reflect fundamental differences in cognitive processing constraints. Writing allows time for deliberate lexical retrieval, conscious morphological manipulation, and orthographic encoding, which enables learners to draw on written vocabulary depth to produce lexically sophisticated texts (Sukying, 2023). Speaking benefits from form–meaning connections that support fluent lexical access, although it occurs under greater time pressure than writing (Sukying, 2025). In contrast, listening imposes the most severe temporal constraints: learners must rapidly recognize fleeting phonological input without visual support or opportunities for reprocessing (Field, 2019). This asymmetry helps explain the minimal contribution of written vocabulary knowledge to listening success, even when that knowledge is extensive and deep. By contrast, phonological vocabulary knowledge is indispensable.

These empirical patterns suggest a refinement to Hulstijn's (2015) **core–peripheral** framework. The framework successfully predicts the hierarchical superiority of vocabulary over metacognition; however, it requires extension to accommodate modality-specific realizations of core components. We propose that phonological vocabulary knowledge serves as the core for listening and speaking, whereas orthographic vocabulary knowledge serves as the core for reading and writing. This refinement preserves the framework's hierarchical insight by distinguishing core linguistic knowledge from peripheral strategic competence and by recognizing internal differentiation within core components that aligns with processing modality. The limited cross-modal transfer observed here indicates that these modality-specific lexicons develop through largely separate learning experiences (reading for orthographic knowledge and listening for phonological knowledge) with only limited automatic interconnection. The skill-specific vocabulary patterns documented across listening (the present study), speaking, and writing (Sukying, 2025) in similar EFL populations provide converging support for this modality-specific reconceptualization of core linguistic knowledge.

The practical implications are substantial for Chinese EFL contexts, where instruction relies predominantly on written presentation and practice (Zhang & Graham, 2020). The results suggest that this approach cultivates knowledge that is poorly suited to the demands of listening. The 24.4% of variance that is uniquely explained by aural measures represents a meaningful portion of listening ability that written-focused instruction does not systematically develop. Students may possess adequate written vocabulary yet fail to recognize the same words in spoken form, which has serious consequences for academic listening and authentic communication. The parallel finding that written vocabulary depth dominates writing performance (Sukying, 2023) while contributing minimally to listening highlights a mismatch between orthography-focused instruction and the phonological access requirements of aural comprehension. Pedagogical approaches should therefore prioritize modality-matched vocabulary development and ensure that learners acquire phonological representations through systematic listening exposure rather than assuming automatic transfer from written vocabulary knowledge.

5.2 Metacognitive Awareness as Peripheral Support

While metacognitive awareness correlated moderately with listening ($\mathbf{r}$ = .430 for total MALQ), its incremental contribution beyond vocabulary was minimal ($\Delta \mathbf{R}^2 = .022$), yielding a small effect size ($\Delta \mathbf{f}^2 = 0.049$) that contrasts sharply with the large effect for vocabulary ($\mathbf{f}^2 = 1.132$). This quantitative pattern supports the **core-peripheral** distinction and helps reconcile conflicting reports about the importance of metacognition in the literature.

Studies reporting strong metacognitive effects often did not control for vocabulary knowledge, which may confound linguistic and strategic contributions. When vocabulary is statistically controlled, as in the present study and in Wallace (2022), the unique contribution of metacognition diminishes substantially. This pattern suggests that some apparent metacognitive effects partly reflect vocabulary knowledge, enabling more effective strategy deployment rather than strategies directly enhancing comprehension. In other words, students with stronger vocabulary can apply metacognitive strategies more successfully because efficient core linguistic processing leaves cognitive resources available for strategic monitoring.

This peripheral status should not be misinterpreted as unimportant. Rather, it reflects the cognitive architecture of listening comprehension, in which bottom-up lexical processing necessarily precedes top-down strategic application. As cognitive load theory predicts (Anderson, 2015), when core linguistic processing remains effortful due to limited vocabulary, few attentional resources are available for metacognitive monitoring and strategic adjustment. Metacognitive strategies cannot compensate for insufficient vocabulary because students cannot strategize effectively around unknown words; however, these strategies can optimize resource allocation and monitoring once core knowledge is sufficiently automatized, thereby reducing processing demands.

The component-level analysis clarifies which metacognitive dimensions are most important. Problem Solving showed the strongest correlation ($\mathbf{r}$ = .430), suggesting that compensatory tactics are most beneficial when lexical access temporarily fails or ambiguity arises. Planning and Evaluation, as well as Directed Attention, showed similar moderate associations (approximately $\mathbf{r}$ = .40), indicating that pre-listening preparation, post-listening reflection, and attentional control each contribute modestly to performance. Person Knowledge (**i.e.**, self-efficacy beliefs and awareness of listening difficulty) showed the weakest positive relationship (**r** = .385), which implies limited benefit without corresponding strategic action.

Most notably, Mental Translation was negatively correlated with listening proficiency (**r** = −.215, **p** = .005), indicating that reliance on L1 translation impairs rather than supports comprehension. Such reliance can create processing bottlenecks: while learners translate one segment into Chinese, they miss subsequent input, which triggers a cascade of comprehension failures. The negative associations between Mental Translation and all other metacognitive strategies (ranging from **r** = −.211 to −.395) further confirm that it functions as a counterproductive coping mechanism distinct from adaptive strategic processes.

These findings position metacognitive instruction as valuable supplementary support rather than a substitute for vocabulary development. Teaching students to abandon Mental Translation and deploy Problem-Solving, Planning, and Evaluation, as well as directed attention, may yield modest but meaningful improvements. The 2.2% increment could determine success or failure in challenging listening situations. However, the small effect size suggests that metacognitive training may not be sufficient to overcome fundamental vocabulary limitations. Pedagogically, listening instruction should therefore prioritize vocabulary development as its core, with metacognitive strategy training introduced as a beneficial enhancement once an adequate lexical foundation is established.

Crucially, the modest contribution of metacognition must be interpreted in light of the intermediate proficiency of our sample (TEM-4 scores 65–75, corresponding to CEFR B1–B2). Recent research indicates that proficiency level moderates the relationship between metacognitive awareness and listening comprehension. The 2.2% incremental variance observed here likely represents an optimal contribution for intermediate learners who possess enough linguistic resources to implement strategies but who have not yet fully automatized their strategic repertoire. Milliner and Dimoski (2021) reported that lower-proficiency learners (CEFR A2) who received process-based metacognitive instruction showed only modest gains in listening self-efficacy, despite improvements in listening proficiency. This suggests that when linguistic resources are severely constrained, learners lack the processing capacity to implement even well-taught strategies effectively. The present findings align with this pattern: less skilled listeners rely more on counterproductive strategies such as Mental Translation, whereas more skilled listeners deploy adaptive strategies more automatically and efficiently.

This proficiency-dependent pattern suggests a threshold model of development. At lower proficiency levels (CEFR A2 and below), metacognitive awareness may contribute even less than the 2.2% observed here, or it may show negative associations when learners rely on counterproductive strategies that they lack the capacity to abandon (Fu et al., 2023; Muhammadpour et al., 2024). At intermediate levels, metacognitive strategies begin to yield measurable benefits as core processing becomes partially automatized, which frees resources for strategic deployment. At advanced levels, where core processing is substantially automatic, metacognitive strategies may assume greater importance because learners possess both the linguistic foundation and the capacity to deploy sophisticated strategies effectively (Du & Man, 2023). Even then, metacognitive awareness remains facilitative rather than foundational: it enhances performance built on solid core knowledge rather than compensating for its absence.

This proficiency-dependent interpretation aligns with our theoretical view, which positions metacognition as peripheral rather than core. Peripheral components enhance performance when core components function adequately; however, they cannot compensate for a lack of core knowledge. Our intermediate-proficiency sample represents a developmental transition point: learners possess adequate vocabulary to support listening (as indicated by the 53.1% of variance explained by vocabulary measures), yet still benefit modestly from strategies that optimize the use of existing resources. Future research should explicitly test the **core-peripheral** distinction across proficiency levels by comparing beginners, intermediate learners, and advanced learners within a single design. Such work would clarify whether the hierarchical relationships observed here are universal or proficiency-specific.

6. Pedagogical Implications, Limitations, and Future Directions

6.1 Implications for L2 Listening Instruction

The pronounced superiority of aural vocabulary (24.4% unique variance) demands fundamental reconsideration of current vocabulary instruction practices in Chinese EFL contexts. Three evidence-based priorities emerge from our findings.

First, modality-matched vocabulary instruction should become standard practice. Systematic training in phonological form recognition—including segmental phoneme discrimination, prosodic pattern awareness, and connected speech processes (assimilation, elision, weak forms)—must precede or at least run in parallel with orthographic learning. Current approaches that introduce vocabulary exclusively through written texts, expecting automatic transfer to listening, leave students with knowledge “trapped” in the wrong format. Explicit instruction linking orthographic and phonological forms, combined with extensive practice recognizing spoken words across varied accents and speech rates, can strengthen cross-modal connections that do not develop automatically.

Second, assessment reform is essential. Listening proficiency assessments and placement decisions should incorporate aural vocabulary measures because traditional written tests systematically underestimate listening-relevant knowledge by 24–39%. Programs that rely exclusively on written vocabulary assessments may inappropriately place students or misjudge their readiness for academic listening demands. Instruments such as A-LEX provide a validated and efficient means to assess phonological vocabulary knowledge.

Third, metacognitive strategy instruction should be positioned as a beneficial supplement rather than a primary intervention. The modest increment (2.2%) confirms that strategy training cannot substitute for vocabulary development. Teachers should prioritize helping students abandon counterproductive mental translation while cultivating adaptive problem-solving and attention management strategies.

However, these metacognitive enhancements will be most effective once students possess adequate core vocabulary to support real-time listening comprehension.

6.2 Limitations and Future Research

Three primary limitations warrant careful consideration. First, our cross-sectional design precludes causal inference about the relationship between vocabulary and listening. Although theoretical rationale and modality-matched measurement support the foundational role of aural vocabulary, the question of directionality remains: Does stronger aural vocabulary drive improvements in listening, or does extensive listening practice enhance phonological vocabulary? Longitudinal research that tracks the co-development of vocabulary and listening over multiple semesters would clarify causal pathways and the optimal sequencing of instruction. Such designs could test whether targeted aural-vocabulary training produces listening gains, or whether extensive listening input is more efficient for developing phonological lexicons.

Second, our intermediate-proficiency sample from one Chinese university limits generalizability. The core–peripheral distinction might vary across proficiency levels and learning contexts. At advanced levels, when core processing is more automatic and consumes fewer cognitive resources, metacognition may assume greater importance than the 2.2% increment suggests. Conversely, at lower proficiency levels with even more limited core knowledge, metacognitive contributions may be negligible. Multi-site studies spanning beginner through advanced proficiency, and comparisons of instructed EFL contexts with immersion ESL environments, would establish the framework's boundaries and developmental trajectory.

Third, we assessed receptive vocabulary exclusively. Productive aural vocabulary, defined as learners' ability to retrieve and mentally generate phonological forms, may show even stronger relationships with listening comprehension. Productive knowledge could support predictive processing during real-time comprehension by enabling learners to anticipate upcoming words based on context. Future research that incorporates productive phonological measures alongside our receptive battery would provide a more complete understanding of the role of aural lexical knowledge.

Additional research directions merit investigation. Studies examining whether explicit grapheme–phoneme instruction strengthens cross-modal lexical mappings could determine whether the dissociation we observed reflects instructional gaps rather than inherent processing constraints. Intervention studies that compare phonological-vocabulary training with metacognitive-strategy training in matched samples would provide more definitive evidence about instructional priorities. Finally, replication across languages with varying orthographic depth (shallow: Spanish; deep: English; non-alphabetic: Chinese) would test whether the modality dissociation patterns observed here reflect universal psycholinguistic constraints or English-specific factors.

Despite these limitations, our findings provide robust evidence that vocabulary knowledge, particularly in its aural, modality-matched form, constitutes the core of L2 listening comprehension, with metacognitive awareness serving as a valuable but peripheral supporting function. These results have important implications for the prioritization of instruction, assessment practices, and future research on L2 listening development.

Acknowledgments

We would like to express our heartfelt to our colleagues and student participants for their invaluable support and assistance with data collection. Most importantly, this research project was financially supported by Mahasarakham University. All the support and assistance are really appreciated.

Authors' contributions

Li Dan and Dr. Apisak Sukying were responsible for study design. Li Dan mainly managed the literature review, data collection, and preliminary data analysis. Dr. Apisak Sukying provided ongoing guidance and consultation throughout the study. Nithipong Yothachai prepared supplementary documents for the research project. All the authors jointly analyzed the data, interpreted and discussed the findings. Li Dan drafted the manuscript. Dr. Apisak Sukying and Nithipong Yothachai revised it. All the authors read and approved the final manuscript.

Funding

This research project was financially supported by Mahasarakham University.

Competing interests

The authors declare that they do not have any conflict of interest.

Informed consent

Obtained.

Ethics approval

The Publication Ethics Committee of the Sciedu Press.

The journal's policies adhere to the Core Practices established by the Committee on Publication Ethics (COPE).

Provenance and peer review

Not commissioned; externally double-blind peer reviewed.

Data availability statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Data sharing statement

No additional data are available.

Open access

Copyrights

References

Anderson, J. R. (2015). Cognitive psychology and its implications (8th ed.). Worth Publishers.

Cheng, J., & Matthews, J. (2018). The relationship between three measures of L2 vocabulary knowledge and L2 listening and reading. *Language Testing, 35*(1), 3-25. https://doi.org/10.1177/0265532216676851

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Du, G., & Man, D. (2023). Mediating effect of metacognition on the relationship between listening self-efficacy and L2 listening comprehension. *The Asia-Pacific Education Researcher, 32*(6), 655-664. https://doi.org/10.1007/s40299-022-00684-z

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*(4), 1149-1160. https://doi.org/10.3758/BRM.41.4.1149

Field, J. (2019). Second language listening: Current ideas, current issues. In J. W. Schwieter & A. Benati (Eds.), *The Cambridge handbook of language learning* (pp. 283–319). Cambridge University Press. https://doi.org/10.1017/9781108333603.013

Fu, Y., Wang, M., Min, S., Zhou, S., & Pan, X. (2023). Exploring the relationship between metacognitive awareness and Chinese EFL learners' listening skills. *Frontiers in Psychology, 14,* Article 1148610. https://doi.org/10.3389/fpsyg.2023.1148610

Ha, H. T. (2021). Exploring the relationships between various dimensions of receptive vocabulary knowledge and L2 listening and reading comprehension. *Language Testing in Asia, 11,* Article 20. https://doi.org/10.1186/s40468-021-00131-8

Hair, J. F., Jr., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th ed.). Cengage Learning.

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

Hulstijn, J. H. (2015). *Language proficiency in native and non-native speakers: Theory and research* (Language Learning & Language Teaching, Vol. 41). John Benjamins Publishing Company. https://doi.org/10.1075/lllt.41

Jin, Y., & Fan, J. (2011). Test for English Majors (TEM) in China. *Language Testing, 28*(4), 589-596. https://doi.org/10.1177/0265532211414852

Keith, T. Z. (2015). *Multiple regression and beyond: An introduction to multiple regression and structural equation modeling* (2nd ed.). Routledge. https://doi.org/10.4324/9781315749099

Masrai, A. (2019). Vocabulary and reading comprehension revisited: Evidence for high-, mid-, and low-frequency vocabulary knowledge. *SAGE Open, 9*(2), Article 2158244019845182. https://doi.org/10.1177/2158244019845182

Masrai, A. (2020). Exploring the impact of individual differences in aural vocabulary knowledge, written vocabulary knowledge and working memory capacity on explaining L2 learners' listening comprehension. *Applied Linguistics Review, 11*(3), 423-447. https://doi.org/10.1515/applirev-2018-0106

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

Matthews, J. (2018). Vocabulary for listening: Emerging evidence for high and mid-frequency vocabulary knowledge. *System, 72,* 23-36. https://doi.org/10.1016/j.system.2017.10.005

Matthews, J., & Cheng, J. (2015). Recognition of high frequency words from speech as a predictor of L2 listening comprehension. *System, 52,* 1-13. https://doi.org/10.1016/j.system.2015.04.015

Milliner, B., & Dimoski, B. (2024). The effects of a metacognitive intervention on lower-proficiency EFL learners' listening comprehension and listening self-efficacy. *Language Teaching Research, 28*(2), 679-713. https://doi.org/10.1177/13621688211004646

Milton, J., & Hopkins, N. (2005). *Aural Lex [Unpublished vocabulary assessment software].* Swansea University.

Milton, J., & Hopkins, N. (2006). Comparing phonological and orthographic vocabulary size: Do vocabulary tests underestimate the knowledge of some learners? *The Canadian Modern Language Review, 63*(1), 127-147. https://doi.org/10.3138/cmlr.63.1.127

Muhammadpour, M., Zafarghandi, A. M., & Tahriri, A. (2024). The effect of metacognitive intervention on the listening performance and metacognitive awareness of high- and low-working memory capacity EFL learners. *Journal of Psycholinguistic Research, 53*(5), Article 70. https://doi.org/10.1007/s10936-024-10102-8

Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781139858656

Nation, I. S. P. (2022). *Learning vocabulary in another language* (3rd ed.). Cambridge University Press. https://doi.org/10.1017/9781009093873

Qian, D. D. (1999). Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension. *Canadian Modern Language Review, 56*(2), 282-308. https://doi.org/10.3138/cmlr.56.2.282

Qian, D. D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language Learning, 52*(3), 513-536. https://doi.org/10.1111/1467-9922.00193

Qian, D. D., & Lin, L. H. F. (2020). The relationship between vocabulary knowledge and language proficiency. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 66-80). Routledge. https://doi.org/10.4324/9780429291586-5

Read, J. (1998). Validating a test to measure depth of vocabulary knowledge. In A. J. Kunnan (Ed.), *Validation in language assessment: Selected papers from the 17th language testing research colloquium, Long Beach* (pp. 41–60). Lawrence Erlbaum Associates.

Schmitt, N. (2014). Size and depth of vocabulary knowledge: What the research shows. *Language Learning, 64*(4), 913-951. https://doi.org/10.1111/lang.12077

Schmitt, N., & Schmitt, D. (2014). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching, 47*(4), 484-503. https://doi.org/10.1017/S0261444812000018

Stæhr, L. S. (2009). Vocabulary knowledge and advanced listening comprehension in English as a foreign language. *Studies in Second Language Acquisition, 31*(4), 577-607. https://doi.org/10.1017/S0272263109990039

Sukying, A. (2023). The role of vocabulary size and depth in predicting postgraduate students' second language writing performance. *LEARN Journal: Language Education and Acquisition Research Network, 16*(1), 575-603. https://so04.tci-thaijo.org/index.php/LEARN/article/view/263457

Sukying, A. (2025). The role of vocabulary knowledge in L2 speaking and writing abilities of Thai EFL undergraduates. *GEMA Online Journal of Language Studies, 25*(3). https://doi.org/10.17576/gema-2025-2503-10

Tabachnick, B. G., & Fidell, L. S. (2019). *Using multivariate statistics* (7th ed.). Pearson.

Taylor, L., & Falvey, P. (Eds.). (2007). *IELTS collected papers: Research in speaking and writing assessment* (Studies in Language Testing, Vol. 19). Cambridge University Press.

Uchihara, T. (2023). How does the test modality of weekly quizzes influence learning the spoken forms of second language vocabulary? *TESOL Quarterly, 57*(2), 595-617. https://doi.org/10.1002/tesq.3176

Vandergrift, L., & Baker, S. (2015). Learner variables in second language listening comprehension: An exploratory path analysis. *Language Learning, 65*(2), 390-416. https://doi.org/10.1111/lang.12105

Vandergrift, L., & Goh, C. C. M. (2012). *Teaching and learning second language listening: Metacognition in action.* Routledge. https://doi.org/10.4324/9780203843376

Vandergrift, L., Goh, C. C. M., Mareschal, C. J., & Tafaghodtari, M. H. (2006). The metacognitive awareness listening questionnaire: Development and validation. *Language Learning, 56*(3), 431-462. https://doi.org/10.1111/j.1467-9922.2006.00373.x

Wallace, M. P. (2022). Individual differences in second language listening: Examining the role of knowledge, metacognitive awareness, memory, and attention. *Language Learning, 72*(1), 5-44. https://doi.org/10.1111/lang.12424

Wang, Y., & Treffers-Daller, J. (2017). Explaining listening comprehension among L2 learners of English: The contribution of general language proficiency, vocabulary knowledge and metacognitive awareness. *System, 65*, 139-150. https://doi.org/10.1016/j.system.2016.12.013

Webb, S., Sasao, Y., & Ballance, O. (2017). The updated Vocabulary Levels Test: Developing and validating two new forms of the VLT. *ITL – International Journal of Applied Linguistics, 168*(1), 33-69. https://doi.org/10.1075/itl.168.1.02web

Zhang, P., & Graham, S. (2020). Learning vocabulary through listening: The role of vocabulary knowledge and listening proficiency. *Language Learning, 70*(4), 1017-1053. https://doi.org/10.1111/lang.12411

Zhang, S., & Zhang, X. (2022). The relationship between vocabulary knowledge and L2 reading/listening comprehension: A meta-analysis. *Language Teaching Research, 26*(4), 696-725. https://doi.org/10.1177/1362168820913998