

INVESTIGATING VOCABULARY RETENTION UNDER SPACED AND MASSED LEARNING CONDITIONS IN THE EFL CLASSROOM

Le Hong Nhung¹, Le Viet Hoai Nam¹, Nguyen Thi Ha¹, Ngo Thi Loan¹, Do Thi Loan¹

Received: 23 April 2026/ Accepted: 20 July 2026/ Published: 30 July 2026

<https://doi.org/10.70117/hdujs.E11.7.2026.1321>

Abstract: *This study examines the effects of spaced and massed practice on vocabulary learning and retention among third-year English majors at Hong Duc University. Grounded in the spacing effect, it compares two conditions: intensive single-session practice and distributed practice across sessions. Participants learned 60 target words, evenly divided between conditions. Vocabulary knowledge was measured through immediate and four-week delayed tests, assessing both recognition and recall. Questionnaire data explored learner perceptions. Results show that spaced practice significantly improved long-term retention, whereas massed practice led to greater decline over time. The advantage of spaced learning was consistent across word types and learner preferences, highlighting its pedagogical value in tertiary EFL contexts.*

Keywords: *EFL learners, massed practice, spaced practice, vocabulary retention.*

1. Introduction

1.1. Background of the study

Vocabulary knowledge is an important part of second language learning and requires repeated exposure for long-term retention [15], [21]. In many EFL contexts, learners often struggle to remember and actively use new vocabulary despite years of formal English instruction. This issue is especially relevant in Vietnam, where English learning mainly takes place in classrooms and opportunities to use English outside school are often limited. Such challenges may be more noticeable among students from rural areas, who may have fewer chances to interact with English in daily life.

Research in cognitive psychology has shown that spaced practice, in which learning is distributed across time, often leads to better long-term memory than massed practice, where learning happens intensively in one session [22], [25]. Spaced learning encourages repeated retrieval and review, which may strengthen memory over time. Although many international studies have supported the spacing effect, its use in authentic EFL classrooms remains underexplored.

In Vietnam, previous research has mainly focused on vocabulary learning strategies, teaching methods, learner autonomy, or general vocabulary retention [18]. Some studies have examined vocabulary size and lexical difficulties among university students, but few have directly compared spaced and massed practice in tertiary EFL settings [19]. In addition, most existing studies rely on short-term tests or focus on only one aspect of vocabulary knowledge. Limited attention has been paid to how different learning schedules affect both recognition and recall among university learners.

¹ Faculty of Foreign Languages, Hong Duc University; Email: lehongnhung@hdu.edu.vn

This study addresses these gaps by examining vocabulary retention under spaced and massed learning conditions among third-year English majors at Hong Duc University. This group was selected because the researcher was directly responsible for teaching these students during the semester, allowing systematic classroom implementation and observation. Although participants had studied English for at least twelve years, their vocabulary size ranged from only 1,000 to 2,100 word families, which may be insufficient for university-level academic reading and language use. Since many students come from rural backgrounds with limited English exposure, investigating effective vocabulary learning schedules in this context is both relevant and necessary.

Specifically, this study addresses the following research questions:

What differences are there between massed and spaced practice in the amount and level of vocabulary knowledge students gain in the immediate post-test?

How do massed and spaced practice compare in terms of students' vocabulary knowledge four weeks after instruction (delayed post-test)?

Which approach leads to better retention of vocabulary knowledge at different strength levels over time?

Do different word classes (nouns vs. verbs) show differences in how well they are learned and retained under massed and spaced conditions?

To what extent do students perceive differences in the effectiveness of spaced practice compared to massed practice?

1.2. Theoretical framework

1.2.1. Instructed vocabulary learning and learnability factors

Vocabulary serves as the fundamental building block of language, requiring systematic pedagogical attention to overcome learners' lexical gaps [14], [16]. In instructed environments, intentional learning, involving deliberate cognitive processes and explicit instruction, is considered indispensable for L2 vocabulary acquisition, as incidental learning often occurs at a slow pace [7], [20]. Furthermore, vocabulary learnability is influenced by various factors, most notably word class. Empirical evidence consistently suggests that nouns are more easily acquired and memorized than verbs, as mapping a word to a concrete object is cognitively less demanding than mapping it to an abstract action [6], [11], [21].

1.2.2. The spacing effect in vocabulary retention

The "spacing effect" posits that information retention is significantly improved when learning episodes are distributed over time (spaced practice) rather than concentrated into a single, continuous session (massed practice) [3], [4]. While massed practice may promote rapid short-term uptake, spaced learning fosters long-term memory traces more effectively. This is supported by theoretical frameworks such as encoding variability (creating richer temporal and cognitive cues) and study-phase retrieval (forcing effortful retrieval of prior memory traces) [8], [13]. Although the superiority of spaced scheduling has been proven in primary EFL classrooms [5], [24], studies at the tertiary level, especially those utilizing multidimensional measures of both recognition and recall remain insufficient, presenting a clear research gap.

1.2.3. Vocabulary testing

Lexical competence is a complex continuum encompassing both breadth (vocabulary size) and depth (receptive vs. productive knowledge) [12], [14]. Since a single test format may fail to capture a learner's true lexical proficiency, this study adopts the implicational hierarchy proposed by [10]. This framework requires applying multiple measures across four distinct levels of vocabulary strength, ordered from least to most demanding: receptive recognition, productive recognition, receptive recall, and productive recall.

2. Methodology

2.1. Research design and participants

This study employed a mixed-methods design from a pragmatist perspective, integrating quantitative and qualitative approaches to generate a comprehensive understanding of vocabulary retention that transcends numerical data alone [2], [23]. The participants were 82 third-year English majors (16 males, 66 females) enrolled in an academic reading and writing module at Hong Duc University, Vietnam. Specifically, the study utilized a within-participant (repeated measures) design, wherein all 82 learners experienced both the spaced and massed learning conditions. To minimize order effects, the two instructional conditions were implemented sequentially across different weeks using two distinct but equivalent sets of target words. Their baseline vocabulary sizes, estimated via the Vocabulary Size Test, ranged from 1,000 to 2,100 word families [17].

2.2. Instruments

Data were collected using three primary instruments:

Questionnaires: A background questionnaire screened for external variables (e.g., overseas experience, health factors) that might influence memory or language exposure. A post-intervention questionnaire, consisting of eight Likert-scale items and one open-ended question, was developed to evaluate learners' perceptions of engagement and learning benefits under both conditions. For internal consistency, the quantitative Likert-scale items indicated strong reliability, yielding a Cronbach's alpha of .84 for the learning construct and .72 for the engagement construct. Regarding the open-ended item intended for thematic qualitative analysis, no responses were obtained from the participants; consequently, analysis was strictly confined to the quantitative data.

Vocabulary Size Test (VST): This multiple-choice instrument [1] was utilized to assess learners' baseline receptive vocabulary across different word-frequency levels. The VST is widely established as a highly reliable tool in SLA research. Scoring was strictly objective, with one point awarded for each correct response, allowing for an estimation of overall vocabulary size.

Vocabulary Achievement Test of Strength (VATS): Adapted to measure knowledge of 60 target words, the VATS assessed four dimensions of vocabulary mastery: productive recall, receptive recall, productive recognition, and receptive recognition [10].

Piloting: Prior to the main intervention, the VATS was piloted with a small cohort of non-participating third-year English majors to verify instruction clarity, time limits, and appropriate difficulty.

Parallel versions: To mitigate test-retest practice effects between the immediate and delayed assessments, two parallel test versions were meticulously developed. The target words remained identical, but the order of items and multiple-choice distractors were randomized using the online Research Randomizer tool.

Scoring criteria: Each of the four subtests was marked out of 30. Scoring was dichotomous: one point was awarded for each correct response, with only a single acceptable answer per item. For productive recall tasks, inaccurate spellings were strictly penalized and treated as incorrect responses.

Sample items: For the receptive recognition level, students selected the correct L1 Vietnamese translation for a given L2 English word (e.g., anomaly). Conversely, for the productive recall level, learners were required to supply the correct L2 English target word when provided with its L1 meaning and the initial letter as a prompt.

2.3. Procedure and data analysis

The intervention was embedded into regular classroom practices over a 10-week period. To control for unintended incidental exposure, three concrete measures were implemented. First, all regular teaching materials were meticulously screened before each lesson, and any occurrence of the target words was replaced. Second, participants were strictly prohibited from keeping learning materials or making notes for out-of-class study during the intervention. Finally, to account for external variables, learners completed a self-report survey following the delayed post-test to document any incidental encounters with the target vocabulary outside the classroom setting.

To ensure comparability, the 60 target words were evenly divided into two sets of 30. Both sets were carefully matched for key psycholinguistic variables, including word class distribution (15 nouns and 15 verbs per condition) and word frequency (all items belonged to the B2-C1 bands on the Academic Word List, verified via the Compleat Lexical Tutor). Furthermore, independent samples t-tests indicated no significant differences between the spaced and massed sets regarding average word length ($M = 8.67$, $SD = 1.54$ vs. $M = 8.70$, $SD = 1.78$, respectively; $p = .94$) and baseline familiarity (both sets yielded a mean pre-test score of 0). For example, Set 1 included items such as “anomaly” and “allocate”, while Set 2 included comparable items like “catalyst” and “accumulate”. This rigorous matching ensures that differences in retention are attributable to the learning conditions rather than intrinsic word characteristics.

In the spaced condition, 30 target words were introduced and subsequently reviewed across four 15-minute sessions distributed over four weeks. In the massed condition, the remaining 30 words were practiced using identical task types, but completed consecutively within a single continuous session.

To evaluate both initial acquisition and retention, immediate post-tests were administered right after the final instruction phase, followed by delayed post-tests exactly four weeks later [19]. Quantitative data from test scores and closed-ended questionnaire items were analyzed using SPSS (Version 21.0). Descriptive statistics were initially generated to summarize central tendencies and assess data distribution. Because skewness and kurtosis values indicated distinct deviations from normality, specifically ceiling effects at the recognition levels and floor effects at the productive recall level, nonparametric tests were required. Consequently, Wilcoxon signed rank tests were selected to compare performance between the spaced and massed conditions, with a Bonferroni adjustment applied to control for Type I errors across multiple comparisons. Furthermore, Generalized Linear Models (GZLM) utilizing Generalized Estimating Equations (GEE) were explicitly chosen to analyze the delayed test results. GZLM is robust against non-normal distributions and is highly effective for evaluating repeated-measures designs with complex interactions (e.g., learning condition $\times$ time $\times$ vocabulary strength). Regarding the qualitative data, the open-ended questionnaire item was designed to be evaluated using thematic qualitative analysis.

3. Findings and discussions

This study contributes to the existing literature in several ways. First, it examines spaced and massed vocabulary learning in an authentic Vietnamese tertiary EFL classroom, a context that has received limited attention in previous research. Second, unlike many earlier studies that focused mainly on a single dimension of vocabulary knowledge, the present study measures vocabulary performance across four levels, including productive recall, receptive recall, productive recognition, and receptive recognition. Third, the combination of immediate and delayed post-tests allows for a clearer examination of both short-term learning and long-term retention. Finally, the study extends previous work by considering word class differences, comparing the retention patterns of nouns and verbs under different learning conditions.

3.1. *Effects of massed vs. spaced practice on initial acquisition*

Descriptive statistics for the immediate post-tests (see Table 1) reveal that spaced practice consistently yielded higher scores than massed practice across all vocabulary strength levels. However, an in-depth analysis indicates a significant interaction between learning condition and vocabulary strength. The advantage of spaced learning was substantially more pronounced in cognitively demanding, open-response tasks, with mean differences reaching 4.4 words for receptive recall and 3.3 words for productive recall. In contrast, differences at the recognition levels were minimal (1.3 for productive and 0.4 for receptive recognition).

Wilcoxon Signed-Rank Tests confirmed that spaced practice produced significantly higher scores at three levels ($p < .004$), but not at the receptive recognition level ($p = .052$). This non-significant result at the lowest difficulty level is primarily attributed to a ceiling effect, as mean scores in both conditions approached the maximum score of 30, thereby restricting score variability and observable differences.

Table 1. Descriptive statistics for immediate post-test scores on all strength levels

		Strength test			
		Productive recall	Receptive recall	Productive recognition	Receptive recognition
Massed	Mean	3.09	11.43	26.66	29.29
	Median	3.01	10.00	27.99	30.00
	SD	2.08	5.30	3.23	1.31
Spaced	Mean	6.38	15.76	27.97	29.69
	Median	6.00	16.00	29.00	30.00
	SD	2.77	5.32	2.53	0.65

3.2. Long-term retention and forgetting patterns

Regarding long-term retention measured four weeks later (see Table 2), a general decline in scores was observed across all strength levels, reflecting natural memory decay. Nevertheless, the forgetting curve was notably steeper for the massed condition.

Table 2. Descriptive statistics for delayed post-test scores on all strength levels

		Strength test			
		Productive recall	Receptive recall	Productive recognition	Receptive recognition
Massed	Mean	0.90	6.85	24.22	27.15
	Median	1.00	6.00	25.00	28.00
	SD	0.96	4.55	3.95	2.71
Spaced	Mean	4.15	14.45	27.64	29.16
	Median	4.00	15.00	29.00	30.00
	SD	2.25	5.53	2.54	1.30

Spaced learning resulted in better long-term vocabulary retention across all strength levels in the delayed post-test (Table 2). The largest difference was found in receptive recall, where learners in the spaced condition achieved a mean score of 14.45 compared to 6.85 in the massed condition, showing a gap of about 7.6 words. Over the four-week period, scores in the massed condition declined more sharply, particularly in receptive recall (from 11.43 to 6.85), while learners in the spaced condition showed a smaller reduction (from 15.76 to 14.45). These findings suggest that spaced practice may support slower forgetting and better retention over time.

A Generalized Linear Model (GZLM) analysis (Table 3) indicated a strong learning condition $\times$ time interaction effect (Wald $\chi^2 = 60.43$, $p < .001$), underscoring differential retention over time.

Table 3. Overall GZLM analysis of immediate and delayed post-test scores, by learning condition and strength level

Sources	Wald Chi-Square	Df	Sig.
Learning condition	70.533	1	< .001
Time	156.342	1	< .001
Strength	286.423	3	< .001
Learning condition * Time	60.432	1	< .001
Learning condition * Strength	14.143	3	.003
Time * Strength	19.675	3	< .001
Learning condition * Time * Strength	6.976	3	.096

Although the three-way interaction among learning condition, time, and vocabulary strength was not statistically significant (Wald $\chi^2 = 6.976$, $p = .096$), this result still provides useful information.

3.3. The impact of word class on vocabulary learnability

Descriptive data for both immediate (Table 4) and delayed post-tests (Table 5) indicate that nouns consistently outperformed verbs across all instructional conditions and assessment times.

Table 4. Descriptive statistics for word class scores in the immediate post-test at four strength levels by learning condition

		Massed				Spaced			
		Productive recall	Receptive recall	Productive recognition	Receptive recognition	Productive recall	Receptive recall	Productive recognition	Receptive recognition
nouns	Mean	2.22	6.75	13.65	14.78	4.12	9.78	14.68	14.89
	Median	2.00	7.00	14.00	15.00	4.00	10.00	15.00	15.00
	SD	1.23	2.78	1.96	0.41	2.11	3.34	0.92	0.34
verbs	Mean	1.21	4.34	12.98	14.67	1.97	6.87	13.85	14.89
	Median	1.00	4.00	14.00	15.00	2.00	7.00	14.00	15.00
	SD	1.32	3.02	2.15	0.98	1.52	3.10	1.47	0.68

Table 5. Descriptive statistics for word class scores in the delayed post-test at four strength levels by learning condition

		Massed				Spaced			
		Productive recall	Receptive recall	Productive recognition	Receptive recognition	Productive recall	Receptive recall	Productive recognition	Receptive recognition
nouns	Mean	1.23	5.78	13.67	14.65	2.92	8.86	14.16	14.87
	Median	2.00	7.00	14.00	15.00	4.00	10.00	15.00	15.00
	SD	1.23	2.78	1.96	0.41	2.11	3.34	0.92	0.34
verbs	Mean	1.21	4.34	12.98	14.67	1.97	6.87	13.85	14.89
	Median	1.00	4.00	14.00	15.00	2.00	7.00	14.00	15.00
	SD	1.32	3.02	2.15	0.98	1.52	3.10	1.47	0.68

Furthermore, a GZLM analysis (Table 6) showed a significant main effect of word class (Wald $\chi^2 \approx 78$, $p < .001$). Crucially, the absence of a significant interaction between word class and learning condition ($p > .05$) suggests that the benefits of spaced practice were relatively similar for both nouns and verbs. Although nouns generally achieved higher scores than verbs, spaced learning appeared to support retention across both word classes.

Table 6. Overall GZLM analysis of immediate and delayed post-test scores, by learning condition, strength and word class

Sources	Wald Chi-Square	Df	Sig.
Time	207.322	1	< .001
Learning condition	78.345	1	< .001
Strength	467.873	3	< .001
Word class	75.322	1	< .001
Time * Learning condition	12.322	1	.001
Time * Strength	32.754	3	< .001
Time * Word class	78.434	1	< .001
Word class * Learning condition	0.325	1	.597
Word class * Strength	14.545	3	.002
Learning condition * Strength	27.392	3	< .001
Time * Word class * Learning condition	0.753	1	.367
Time * Word class * Strength	13.003	3	.005
Word class * Learning condition * Strength	3.163	3	.367
Time * Word class * Learning condition * Strength	7.559	3	.056

However, a significant interaction between word class and time (Wald $\chi^2 = 78.434$, $p < .001$), together with the Time $\times$ Word class $\times$ Strength interaction ($p = .005$), suggests that vocabulary decay patterns differed across word types and knowledge levels. Verbs generally showed greater loss over time than nouns, particularly at the receptive recall level. One possible explanation is that verbs are often semantically and grammatically more complex than nouns, which may increase learning difficulty and reduce retention. By contrast, nouns may benefit from clearer conceptual meaning and stronger form-meaning associations. However, this interpretation should be treated cautiously, as the present study focused on a limited number of target words.

3.4. Learners' perceptions and engagement

Analysis of the post-intervention questionnaire provided nuanced insights into learner psychology. Overall mean ratings clustered around neutrality ($M \approx 2.31$ - 3.72) (Table 7), suggesting that students did not possess an overwhelming explicit preference for either practice type.

Table 7. Descriptive statistics for questionnaire responses concerning learners' perceptions of learning conditions

Construct	Items	Mean	Median	SD
Learning	a. I memorize words better in spaced learning than massed learning.	2.98	3	0.95
	b. I memorize more words in spaced learning than massed learning.	2.76	3	0.69
	c. I retain words better in spaced learning than massed learning.	3.02	3	0.84
	d. I recall words better in spaced learning than massed learning.	3.32	3	0.82
	e. I learn words more quickly in spaced learning than massed learning.	2.89	3	0.62
Engagement	f. I can focus better in spaced learning than massed learning.	2.31	3	0.77
	g. I feel more motivated in spaced learning than massed learning.	3.27	3	0.68
	h. I feel less bored in spaced learning than massed learning	3.72	3	0.83

Breaking down the constructs, learners recognized the cognitive advantages of spaced learning for word recall ($M = 3.32$) and reported feeling significantly less bored ($M = 3.72$) and more motivated ($M = 3.27$) under the spaced condition.

3.5. Discussion

3.5.1. Interpretation of vocabulary retention and memory mechanisms

One possible explanation for the long-term retention patterns comes from memory theory. Spaced learning requires learners to retrieve vocabulary after intervals, which may

strengthen memory through effortful retrieval. Repeated encounters across time may also help learners form richer memory connections. In contrast, massed practice may support short-term familiarity but appears less effective for long-term retention. In both conditions, recognition scores remained high (above 24), whereas recall scores were considerably lower, suggesting that recognizing vocabulary is cognitively less demanding than retrieving it independently.

Furthermore, the non-significant three-way interaction observed in the GZLM analysis suggests that the advantage of spaced practice remained relatively consistent across different vocabulary strength levels over time. In other words, while the size of the spacing effect varied slightly depending on task difficulty, there was insufficient evidence to conclude that forgetting patterns differed substantially across vocabulary strength categories. This finding indicates that spaced practice benefits multiple dimensions of vocabulary knowledge rather than only a particular level of mastery.

3.5.2. The role of word class and learner perceptions

Consistent with prior literature, descriptive data indicated that nouns consistently outperformed verbs across all instructional conditions and assessment times. This finding aligns with the cognitive load framework, which suggests that mapping a word to a concrete object (a noun) is cognitively less demanding than mapping it to an abstract action or mental state (a verb). Crucially, the absence of a significant interaction between word class and learning condition implies that the inherent learnability differences between parts of speech operate independently of instructional scheduling; the spacing effect remains universally beneficial for both word categories.

Regarding learner subjective experiences, while the descriptive data showed that spaced practice reduced boredom and enhanced motivation, the actual preference ratings remained relatively neutral. As noted in the questionnaire responses, moderate variability and the fact that a subset of learners still explicitly favored massed practice highlight the strong influence of individual learning styles and the deeply entrenched habit of cramming in EFL contexts.

3.5.3. Pedagogical implications

The empirical evidence obtained from this study offers several actionable implications for tertiary EFL instruction. First, because massed practice leads to rapid long-term memory decay, curriculum designers and instructors should move away from intensive, single-session vocabulary drills. Instead, vocabulary recycling should be systematically distributed across multiple weeks. Second, given the substantial gap observed between high recognition scores and low recall scores, classroom assessments and activities must go beyond simple multiple-choice formats. Teachers should incorporate more production-oriented tasks, such as sentence generation and gap-filling, to foster deeper vocabulary retrieval. Finally, since students are highly accustomed to massed cramming, explicit strategy training should be provided to help learners recognize and apply the cognitive benefits of spaced self-study.

4. Conclusion

This study investigated the effects of spaced versus massed practice on vocabulary learning and retention among university EFL learners. Contributing to the literature by

taking place in an authentic classroom setting, the research examined how instructional timing interacts with word class and varying depths of vocabulary knowledge over time. Participants acquired target vocabulary through structured exercises delivered either in a single, concentrated session (massed condition) or distributed over four weeks (spaced condition), with learning outcomes comprehensively assessed via immediate and delayed multidimensional tests.

Findings firmly establish that spaced practice yields greater initial learning gains and significantly superior long-term retention compared to massed practice. While nouns remained generally easier to acquire than verbs, the spacing effect indicated relatively beneficial across word classes and regardless of learners' explicit instructional preferences. Given the fundamental role of lexical competence in language development, these results strongly advocate for the systematic integration of spaced repetition into tertiary EFL curricula. Future research should explore the optimal practical implementation of this approach and determine which agents - teachers, materials, or learners themselves - play the most effective role in driving it.

References

- [1] Beglar, D. (2010), *A Rasch-based validation of the Vocabulary Size Test*. Language Testing, 27(1):101-118. <https://doi.org/10.1177/0265532209340194>
- [2] Biesta, G. (2021). Mixing methods in educational research. In R. Coe, M. Waring, L. V. Hedges, & L. Day Ashley (Eds.), *Research methods and methodologies in education* (3rd ed., pp. 186-193). SAGE Publications.
- [3] Dempster, F. N. (1987), *Effects of variable encoding and spaced presentations on vocabulary learning*. Journal of Educational Psychology, 79(2):162-170. <https://doi.org/10.1037/0022-0663.79.2.162>
- [4] Ebbinghaus, H. (1964), *Memory: A contribution to experimental psychology* (H. A. Ruger, C. E. Bussenius, & E. R. Hilgard, Trans.). Dover Publications. (Original work published 1885).
- [5] Goossens, N. A., Camp, G., Verkoeijen, P. P., Tabbers, H. K., & Zwaan, R. A. (2012), *Spreading the words: A spacing effect in vocabulary learning*. Journal of Cognitive Psychology, 24(8):965-971.
- [6] Horst, M., & Meara, P. (1999), *Test of a model for predicting second language lexical growth through reading*. Canadian Modern Language Review, 56(2):308-328. <https://doi.org/10.3138/cmlr.56.2.308>
- [7] Hulstijn, J. H. (2003), *Incidental and intentional learning*. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 349-381). Blackwell.
- [8] Kapler, I. V., Weston, T., & Wiseheart, M. (2015), *Spacing in a simulated undergraduate classroom: Long-term benefits for factual and higher-level learning*. Learning and Instruction, 36:38-45. <https://doi.org/10.1016/j.learninstruc.2014.11.001>
- [9] Laufer, B. (1997), *What's in a word that makes it hard or easy? Intralexical factors affecting the difficulty of vocabulary acquisition*. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 140-155). Cambridge University Press.
- [10] Laufer, B., & Goldstein, Z. (2004), *Testing vocabulary knowledge: Size, strength, and computer adaptiveness*. Language Learning, 54(3):399-436.

- [11] Maguire, M. J., Hirsh-Pasek, K., & Golinkoff, R. M. (2006), *A unified theory of word learning: Putting verb acquisition in context*. In K. Hirsh-Pasek & R. M. Golinkoff (Eds.), *Action meets word: How children learn verbs* (pp. 364-391). Oxford University Press.
- [12] Melka, F. (1997), *Receptive vs. productive aspects of vocabulary*. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 84-102). Cambridge University Press.
- [13] Melton, A. W. (1970), *The situation with respect to the spacing of repetitions and memory*. *Journal of Verbal Learning and Verbal Behavior*, 9(5):596-606.
- [14] Milton, J. (2009), *Measuring second language vocabulary acquisition* (Vol. 24). *Multilingual Matters*.
- [15] Nation, I. S. P. (2001), *Learning vocabulary in another language*. Cambridge University Press.
- [16] Nation, I. S. P. (2006), *How large a vocabulary is needed for reading and listening?* *Canadian Modern Language Review*, 63(1):59-82. <https://doi.org/10.3138/cmlr.63.1.59>
- [17] Nation, I. S. P., & Beglar, D. (2007), *A vocabulary size test*. *The Language Teacher*, 31(7):9-13.
- [18] Ngo, T. C. T., & Doan, L. A. T. (2023), *EFL students' strategies for learning academic vocabulary with mobile phones at Van Lang University in Vietnam*. *International Journal of Language Instruction*, 2(4).
- [19] Pham, T. T., Nguyen, L. A. D., Dang, H. M., & Le, T. T. P. (2023), *Exploring tertiary Vietnamese EFL students' engagement in vocabulary learning through the use of an AI tool*. In *Proceedings of the AsiaCALL International Conference* (Vol. 4, pp.129-149). Atlantis Press.
- [20] Schmitt, N. (2008), *Review article: Instructed second language vocabulary learning*. *Language Teaching Research*, 12(3):329-363. <https://doi.org/10.1177/1362168808089921>
- [21] Schmitt, N. (2010), *Researching vocabulary: A vocabulary research manual*. Palgrave Macmillan.
- [22] Seabrook, R., Brown, G. D., & Solity, J. E. (2005), *Distributed and massed practice: From laboratory to classroom*. *Applied Cognitive Psychology*, 19(1):107-122. <https://doi.org/10.1002/acp.1066>
- [23] Sharma, L. R., Bidari, S., Bidari, D., Neupane, S., & Sapkota, R. (2023), *Exploring the mixed methods research design: Types, purposes, strengths, challenges, and criticisms*. *Global Academic Journal of Linguistics and Literature*, 5(1):3-12. <https://doi.org/10.36348/gajll.2023.v05i01.002>
- [24] Sobel, H. S., Cepeda, N. J., & Kapler, I. V. (2011), *Spacing effects in real-world classroom vocabulary learning*. *Applied Cognitive Psychology*, 25(5):763-767. <https://doi.org/10.1002/acp.1747>
- [25] Toppino, T. C., & Bloom, L. C. (2002), *The spacing effect, free recall, and two-process theory: A closer look*. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(3):437-444. <https://doi.org/10.1037/0278-7393.28.3.437>