International Journal of Language and Literary Studies
Volume 8, Issue 5, 2026
Homepage : http://ijlls.org/index.php/ijlls

Incidental Vocabulary Learning: Evaluating the Predictions of the Involvement Load Hypothesis Plus

Maram S. Almohaimeed
English Language Department, College of Language Sciences, King Saud University, Saudi Arabia
ORCID: https://orcid.org/0000-0002-0275-3239
Corresponding Author Email: msmohaimeed@ksu.edu.sa

DOI: https://doi.org/10.36892/ijlls.v8i5.2757

APA Citation: Almohaimeed, M. S. (2026). Incidental Vocabulary Learning: Evaluating the Predictions of the Involvement Load Hypothesis Plus. *International Journal of Language and Literary Studies*, *8*(5).247-262. https://doi.org/10.36892/ijlls.v8i5.2757

Received: *15/05/2026*

Accepted: *25/07/2026*

Keywords: *incidental vocabulary learning, Involvement Load Hypothesis Plus, search, word length, working memory.*

Abstract

Incidental vocabulary learning is a fundamental way to develop L2 learners' vocabulary knowledge. The Involvement Load Hypothesis Plus (ILH Plus) offers a refined model of the earlier ILH on incidental vocabulary learning, yet its predictions remain largely unverified. This study evaluated the predictions of the ILH Plus regarding the contribution of the search component and examined whether word length and working memory (WM) contribute beyond the model or moderate its predictions. In a quasi-experimental design, 87 female university EFL learners completed a reading task either with or without search (dictionary consultation). Vocabulary learning was assessed using immediate and delayed meaning-recall tests, and WM was measured using digit-span tasks. Results showed that both tasks produced comparable immediate vocabulary learning which supports the predictions of the ILH Plus whereas similar delayed retention does not. Word length significantly predicted both immediate and delayed vocabulary learning whereas phonological WM was only positively associated with immediate learning. Neither word length nor WM moderated the predictions of the ILH Plus. The findings suggest that there is a threshold for task effectiveness to be measurable when compared to another task. They also emphasize the importance of considering lexical characteristics alongside task effectiveness index when evaluating incidental vocabulary learning.

1. Introduction

Incidental vocabulary learning is a major source of second language (L2) vocabulary development (Webb, 2019). As learners engage in meaning-focused activities such as reading, they occasionally acquire new words without trying to memorize them deliberately. Because vocabulary knowledge is important to language proficiency, identifying the factors that facilitate incidental vocabulary learning remains an important endeavour of vocabulary research.

Among the theoretical frameworks proposed to explain task-induced vocabulary learning, the Involvement Load Hypothesis (ILH) (Laufer & Hulstijn, 2001) has received considerable attention. The hypothesis predicts that tasks inducing greater involvement lead to greater vocabulary acquisition and retention. Although many studies supported its predictions (e.g., Kim, 2008; Keating, 2008; Eckerth & Tavakoli, 2012), other studies reported findings that contradict or support the model only partially (e.g., Folse, 2006; Snoder, 2017; Huang, 2018). Drawing on this body of research, Yanagisawa and Webb (2022) developed the ILH Plus as a revised framework that offers more specific predictions regarding the contribution of factors related to task characteristics to immediate vocabulary learning and delayed retention.

The ILH Plus has hardly been empirically evaluated despite its theoretical importance. To date and as far as the researcher knows, only one published study by Komuro (2023) directly tested the framework. Furthermore, Yanagisawa and Webb (2022) invited future research to test their model and to examine whether variables outside the model, such as learner- and word-related characteristics, contribute to vocabulary learning beyond its predictions. This recommendation is particularly important because previous research showed that word length as a lexical characteristic and working memory (WM) as a learner characteristic may affect vocabulary learning. However, these variables have not been investigated within the framework of the ILH Plus.

The present study responds to these gaps by evaluating the predictions of the ILH Plus regarding the search component while examining whether word length and WM contribute beyond the model and moderate its predictions. By integrating task, lexical, and learner characteristics within one single study, this research provides one of the first empirical evaluations of the ILH Plus. Accordingly, the study addresses the following research questions:

1. To what extent does the ILH Plus accurately predict vocabulary learning when the search component is manipulated?
2. Does word length contribute to incidental vocabulary learning, and does it moderate its predictions?
3. Does working memory contribute to incidental vocabulary learning, and does it moderate its predictions?

2. Theoretical Background

The ILH (Laufer & Hulstijn, 2001) was proposed to explain why some vocabulary learning tasks lead to better incidental vocabulary learning than others. Drawing on the Depth of Processing framework (Craik & Lockhart, 1972), the ILH maintains that the effectiveness of a task depends on the amount of involvement it induces during learning. This involvement is determined by three components: need, search, and evaluation. Need refers to the motivation to learn a word while search refers to consulting a source to find out the meaning of an unfamiliar L2 word or the L2 form equivalent to an L1 concept. Evaluation requires learners to compare or use lexical items in context. Each component is assigned an involvement degree, and the sum of these degrees represents the overall involvement load (IL) of a task. The ILH predicts that tasks with higher ILs result in better vocabulary acquisition and retention than tasks with lower ILs.

Since it was introduced, the ILH has generated a large amount of research and has become one of the most investigated frameworks in vocabulary learning (Yanagisawa, 2020). Many studies supported its predictions (e.g., Kim, 2008; Keating, 2008; Eckerth & Tavakoli, 2012). However, other studies reported findings that contradict its predictions or support the hypothesis only partially (e.g., Folse, 2006; Snoder, 2017; Huang, 2018). These findings collectively suggest that although the overall IL is an important predictor of incidental vocabulary learning, the relative contribution of individual involvement components may not be captured by the original model adequately. Other limitations were also identified in recent meta-analyses and reviews such as Hazrat and Read (2022), Liu and Reynolds (2022) and Yanagisawa and Webb (2021). To address these limitations, Yanagisawa and Webb (2022) proposed the ILH Plus by using model-selection procedures and evidence synthesized from their large-scale meta-analysis (Yanagisawa & Webb, 2021). The ILH Plus was introduced as a revised framework of the original ILH for predicting incidental vocabulary learning. Rather than assuming that all involvement components contribute equally, the ILH Plus assigns different weights to individual task characteristics and distinguishes between immediate vocabulary acquisition and delayed retention by proposing separate prediction formulas for each outcome. These formulas are presented below.

The incidental vocabulary learning formula of activities for immediate learning

= [Need (absent : 0 or present : 1) × 20.9]
+ [Evaluation (0 or 1) × 8.3]
+ [Sentence-level varied use (0 or 1) × 15.3]
+ [Composition-level varied use (0 or 1) × 23.3]
+ [Frequency (number of times of encounter or use) × 9.4]
+ [Mode (written : 0 or spoken : 1) × – 9.8]

The incidental vocabulary learning formula of activities for retention

= [Need (absent : 0 or present : 1) × 13.8]
+ [Search (0 or 1) × – 5.1]
+ [Evaluation (0 or 1) × 9.1]
+ [Sentence-level varied use (0 or 1) × 11.5]
+ [Composition-level varied use (0 or 1) × 21.0]

Compared to the original ILH, this framework introduces several important changes. It distinguishes different levels of evaluation and incorporates frequency of encounter/use and mode of activity as additional task characteristics. It also assigns different weights to individual predictors according to their estimated contribution to learning. Most importantly for the present study, the ILH Plus revised the role of search. Whereas the original ILH assumes that searching for the meaning of unfamiliar words facilitates vocabulary learning, the ILH Plus predicts that search does not contribute to immediate vocabulary acquisition and may have a small negative effect on delayed retention. Consequently, when two tasks are identical in all other respects, the model predicts comparable immediate learning but lower delayed retention for the task requiring search.

The ILH Plus has two central predictions. First, when variables included in the model are held constant, tasks with higher effectiveness indices should produce greater vocabulary learning than tasks with lower effectiveness indices. Second, even when variables outside the model differ across tasks, those tasks with higher effectiveness indices should still lead to superior learning outcomes. Despite these theoretically important predictions, empirical evidence evaluating the ILH Plus remains scarce. To the best of the researcher's knowledge, Komuro (2023) is the only study that examined the framework. The study supported the predictions of the ILH Plus regarding the revised evaluation component and the role of learner proficiency, but several predictions of the model remain untested, including those concerning the search component. Additionally, Yanagisawa and Webb (2022) invited future research not only to verify the predictions of the ILH Plus but also to investigate variables beyond task characteristics to improve the explanatory power of the model. Hence the present study aimed to examine the role of search and two external variables, word length and WM capacity. The following sections review the empirical literature related to these variables.

3. Empirical Evidence

Empirical research studying the contribution of search to incidental vocabulary learning is mixed. Studies comparing tasks that involve search to tasks that did not generally reported little or no advantage for search in either immediate or delayed vocabulary learning (e.g., Snoder, 2017; Huang, 2018). In contrast, other studies showed that search operationalized in dictionary consultation can facilitate learning. For example, Yang and Cao (2021) showed superior delayed retention for learners who completed a task with search. Udry and Berthele (2025) found that online dictionary led to greater learning than on-screen glossary and paper dictionary. A similar conclusion was reached by Lo (2024) who found that the online dictionary and sentence-level searches using Google Translate promote vocabulary acquisition and retention. Advantages were also reported for electronic dictionaries compared to paper dictionaries (Dziemianko, 2010, 2017). These findings suggest that the effectiveness of search may depend on the learning conditions and task demands, or even differences in learners' proficiency (Lo, 2024; 2025). However, none of these studies examined the specific prediction

of the ILH Plus, search should not improve immediate vocabulary learning and may even reduce delayed retention when other variables of the model are held constant.

Another variable of interest is word length. It is a lexical characteristic that may influence vocabulary learning, yet it has not been thoroughly examined in L2 vocabulary research (Peters, 2020). Word length has been measured in different ways, as the number of letters, phonemes, or syllables. Generally, longer words are assumed to impose greater processing and memory demands than shorter words, and this may increase learning difficulty. However, the extent to which word length affects vocabulary acquisition is still uncertain because empirical findings show considerable variation across studies. In several studies word length was reported as a significant predictor of vocabulary learning, with shorter words learned or retained more successfully than longer ones (e.g., Alsaif & Milton, 2012; Puimège and Peters, 2019; Reynolds et al., 2015; Vanbuel, 2025; Willis & Ohashi, 2012). However, other studies found that the effect became weak or non-significant after controlling for other lexical variables and learner characteristics (Masrai & Milton, 2015; Zeng et al., 2022; Mulder et al., 2019; Urdaniz & Skoufaki, 2022). These inconsistent findings can be attributed to factors such as learner proficiency, morphological awareness, and differences in research design. Importantly, most studies on word-length effect are correlational or predictive, examining the correlation based on vocabulary knowledge tests rather than employing a task-based model of incidental vocabulary learning. Therefore, whether word length contributes beyond the predictions of the ILH Plus remains unknown.

WM is another potential variable that contributes to vocabulary learning (Baddeley, 2003; Baddeley, 2015; Juffs & Harrington, 2011). It is a system with limited capacity responsible for the temporary storage and processing of information. According to Baddeley and Hitch's (1974) multicomponent model, WM comprises the phonological loop and the central executive alongside other subsystems. The former is responsible for storing verbal information temporarily while the latter controls attention and coordinates cognitive processing. In language learning, both components are relevant because the process requires learners to encode, maintain, and manipulate unfamiliar lexical information. Although WM was examined as a source of individual differences in vocabulary acquisition, its contribution is inconsistent across studies. While several studies reported positive relationships between WM and incidental vocabulary learning (Lim & Lee, 2025; Montero Perez, 2020; Teng, 2024, 2025; Yang et al., 2017), others found that its contribution was limited (Malone, 2018) or absent (Jung et al., 2025; Suárez et al., 2021). Even among studies reporting significant effects of WM, there is little agreement regarding which component of WM, phonological or executive, plays the more important role. These differences were explained in relation to variation in learner proficiency, task demands, assessment of word knowledge, and the measures used to assess WM capacity. Consequently, although WM is widely regarded as an important source of individual differences in vocabulary learning, its contribution within the framework of the ILH Plus has yet to be examined.

4. Methodology

The current study adopted a quantitative quasi-experimental design and employed an immediate-delayed posttest design. Participants completed a reading task followed immediately by a vocabulary meaning-recall test. A delayed posttest was then delivered two weeks later to assess retention. Pretesting to ensure participants' unfamiliarity with the target words was not used since this may result in learning effect.

4.1 Participants

The participant were 87 learners sampled conveniently for administrative constrains. The participants were female low-intermediate EFL learners studying in a Saudi university. Their proficiency level was determined based on their performance in a placement test administrated by the institution upon admission. Participants belong to intact classes which were randomly assigned to either experimental or control group.

4.2 Reading Tasks

Two reading tasks were compared in the experiment: a search task (ST) and a non-search task (NST). The two tasks were almost identical except in the search component. Each task required participants to read a narrative passage of about 150 words, six of which were the target words.

For the ST, the target words were removed from the reading material, and participants had to look them up in the online Cambridge English Dictionary to fill in the gaps. To ensure that participants looked up definitions in the dictionary, they had to write them down on the worksheet before moving on to engage in reading comprehension. After reading the text, participants were given a True/False comprehension exerciseto complete.

In the NST, the same text was utilized with the target items in place. The text was supplemented with a marginal glossary providing English definitions of the target items. The provision of the glossary eliminated the need to use the dictionary. After they finish reading, participants completed a fill-in-the-blanks activity using the target words and a short True/False exercise. The ST and NST were devised so that all other variables specified in the ILH Plus model were held constant. According to the formulas of the model, both tasks had identical effectiveness indices for immediate learning (48) and were therefore expected to yield similar vocabulary acquisition. For retention, the NST received a higher effectiveness index (22.9) than the ST (17.8) and hence expected to lead to different retention.

4.3 Target words

Six target words were selected, and these were: vex, dusk, prowess, absolve, quandary, and permeate. The selection was based on several criteria. The words were low-frequency items according to the British National Corpus and the Corpus of Contemporary American English, as accessed through https://textinspector.com/. To examine word-length effects, the set included two one-syllable, two two-syllable, and two three-syllable words, and these will be referred to as short, medium, and long respectively throughout the remaining of the paper. The target words had no prefixes or common suffixes to avoid the effect of morphological awareness. Also, only words with one or two dictionary senses were selected to reduce the effect of low dictionary-skills reported in Chen (2017). Finally, to control for grammatical class, each word-length category contained one noun and one verb.

Since no pretest was administered, unfamiliarity with the target words was ensured through pilot testing with learners from the same population and scrutiny of course materials the participants were studying. An additional check for unfamiliarity was through including a self-report question in the immediate posttest.

4.4 Measures

Acquisition and retention of the target words were assessed through meaning-recall posttests. In these tests, participants were provided with the target items and were required to provide their meaning in English or Arabic (the learners' L1). A meaning-recall test was preferred to other test formats because it is widely used in incidental vocabulary learning research and more specifically in ILH studies (e.g., Hulstijn & Laufer, 2001). Furthermore, this test format corresponds to the activities of the reading task which required participants to recognize meanings and establish a form-meaning connection. To avoid order effects resulting from repeated testing, the order of the target words was altered in the delayed posttest.

Due to inconsistency in the literature regarding which WM component contributes to vocabulary learning, both phonological and executive WM were examined. Given that digit-span tasks depend only on digits as testing items, these measures were preferred to others that use linguistic stimuli and may be confounded by learners' proficiency (Juffs & Harrington, 2011; Linck et al., 2014). Participants completed a forward-digit span task to measure the phonological WM followed by a backward digit-span task to measure the executive WM. The former task requires information storage while the latter requires information storage and manipulation (Waters & Caplan, 2003). Both tasks were administered using Psytoolkit, a free

web-based platform for cognitive and psychological experiments and surveys (Stoet, 2010; 2017) In the forward digit-span task, participants need to recall in the same order a sequence of digits that increases by one digit after each two successful trials. When a participant failed to recall a sequence for two consecutive trials, the task terminates and the longest digit sequence the participant was able to recall represented the WM span. The backward digit-span measure has the exact same procedure as the forward one except that participants recall the digit sequences in the reverse order. Sequences begin with two digits and can extend up to a maximum of nine digits.

4. 5 Procedure

After obtaining all necessary institutional approvals, data were collected during regular class sessions. Participants first signed an informed consent form and were briefed on the overall aim of the study while not disclosing the specific aim to ensure that they do not engage in deliberate vocabulary learning. Participants then completed the forward digit-span task followed by the backward digit-span task. Next, they completed the reading task. Only those in the ST group were permitted to use their phones to access the online dictionary while the NST group completed the task without dictionary consultation. Immediately after completing the task, participants were given the immediate posttest. Two weeks later and without prior notice, the delayed posttest was administered. All tasks and tests were completed under the researcher's supervision.

4.6 Data Analysis

Responses in the immediate and delayed posttest represented the vocabulary acquisition and retention. Following Hulstijn and Laufer (2001) scoring, responses received 0, 0.5, or 1 for no or inaccurate meaning, semantically similar meaning, or accurate meaning recall, respectively. To establish scoring reliability, the tests were scored by the researcher twice with three-week interval and a high intra-rater reliability was achieved (weighted $\kappa = 0.93$). In addition, 25% of the responses were scored by a colleague English teacher, and a high inter-rater reliability was achieved (weighted $\kappa = .915$).

All statistical analyses were conducted in R. Descriptive statistics were computed for each group and word-length category. Cumulative link mixed models (CLMMs) were fitted using the ordinal package to examine the effects of the independent variables, which were task condition, word length, and the two WM measures, and their interactions. Separate models were fitted for immediate vocabulary learning and delayed retention.

Because the outcome variable consisted of ordered recall scores (0, 0.5, and 1) and each participant contributed multiple observations (six target items), CLMMs were used. CLMMs are appropriate for ordinal response variables while accounting for the non-independence of repeated observations through random effects. This approach also allowed the effects of the independent variables and their interactions to be examined simultaneously.

5. Results and Discussion

The following analyses evaluated the predictive validity of the ILH Plus framework. CLMMs were estimated to examine whether the framework accurately predicted vocabulary learning when the search component was manipulated and whether word length and WM contributed to vocabulary learning. Another set of models were also estimated to examine whether word length and WM moderated the predictions of the ILH Plus. Descriptive statistics are presented first, followed by the CLMM analyses and discussion of the results organised according to the research questions.

5.1 Descriptive Statistics

Table 1 summarizes participants' vocabulary performance across task conditions on immediate and delayed posttests. As shown in Table 1, the ST group achieved slightly higher mean recall score than the NST group on the immediate posttest (0.53 vs. 0.45). In contrast, the NST group showed slightly higher delayed retention than the ST group (0.26 vs. 0.19).

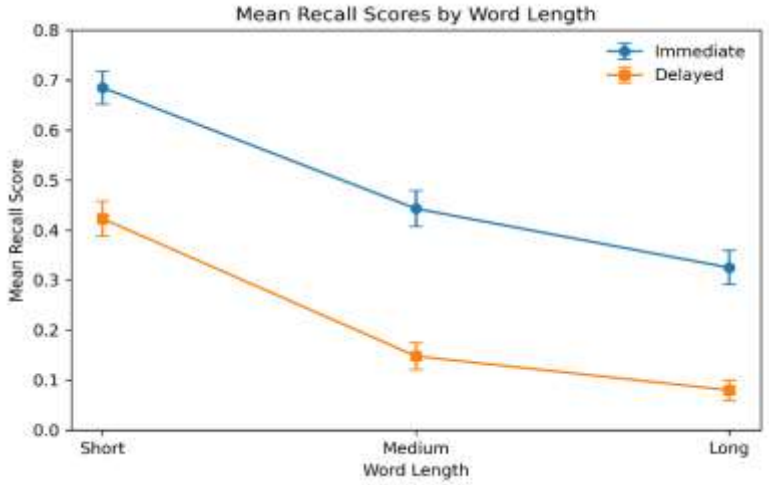
Table 1. Mean Scores by Task Condition at Immediate and Delayed Posttests

Task condition	Immediate – n	Immediate – Mean	Immediate – SD	Delayed – n	Delayed – Mean	Delayed – SD
NST	39	0.44	0.20	39	0.25	0.20
ST	48	0.52	0.25	48	0.19	0.15

Note. Mean recall scores represent participant-level averages across the six target words. Mean score ranges between 0 to 1. n represents the number of participants

Figure 1 presents the mean scores by word-length categories. It can be seen that as word length increased, the recall mean score decreased consistently at both posttests. Short words produced the highest recall scores on both the immediate (M = 0.69, SD = 0.42) and delayed posttests (M = 0.44, SD = 0.45), whereas long words produced the lowest recall scores (Immediate: M = 0.33, SD = 0.46; Delayed: M = 0.08, SD = 0.27). These descriptive patterns were then examined using the CLMMs.

Figure 1. Mean recall scores by word length at the immediate and delayed posttests.



Note. Error bars represent ±1 standard error.

5.2 To what extent does the ILH Plus accurately predict the effects of the search component on incidental vocabulary acquisition and retention?

To address this research question, the estimated CLMM for immediate recall presented in Table 2 showed that task condition did not predict vocabulary performance significantly (β = 0.44, SE = 0.247, z = 1.79, p = .073) although participants in the ST condition obtained numerically higher immediate mean score than those in the NST condition.

Table 2. CLMMs Predicting Main Effects on Immediate and Delayed Recall

Predictor	Immediate Recall – β (SE)	Immediate Recall – z	Immediate Recall – p	Delayed Recall – β (SE)	Delayed Recall – z	Delayed Recall – p
Search (ST)	0.443 (0.247)	1.79	.073	-0.443 (0.300)	-1.47	.141
Medium word	-1.097 (0.291)	-3.77	< .001	-1.677 (0.655)	-2.56	.010
Long word	-1.645 (0.300)	-5.49	< .001	-2.471 (0.685)	-3.61	< .001
Phonological WM	0.265 (0.136)	1.95	.051	0.172 (0.163)	1.06	.290
Executive WM	0.094 (0.123)	0.77	.444	0.065 (0.147)	0.45	.656
Random intercept						
Participant	0.435 (0.660)			0.459 (0.677)		
Word	0.031 (0.175)			0.343 (0.585)		

A similar pattern was observed for delayed retention. Task condition did not significantly predict delayed vocabulary performance ($\beta = -0.44$, SE = 0.30, z = -1.47, p = .141) even though the NST group showed numerically higher delayed mean score than the ST group.

Although descriptive statistics showed numerical advantage for the ST group in immediate learning and for the NST in delayed retention, this advantage was not statistically significant. The absence of difference between the two task conditions in immediate learning support the ILH Plus, which predicts that the two tasks should lead to similar vocabulary gains because they had identical effectiveness indices. The absence of difference between the two conditions in delayed retention, however, ran counter to the predictions of the framework which predicts that the NST should lead to greater learning outcomes since it had a higher effectiveness index. Nevertheless, the retention results should not be interpreted as evidence against the framework because Yanagisawa and Webb (2022) acknowledged that when the difference in the effectiveness indices of two tasks is relatively small, differences in learning outcomes may not be observable. In the present study, the effectiveness indices of the ST and NST for retention were modestly different (17.8 vs. 22.9). Therefore, the absence of difference between the two conditions in retention should not be understood as contradicting the framework. With the numerical advantage of the NST group in retention and Yanagisawa and Webb's (2022) acknowledgement, the findings suggest that differences in predicted task effectiveness may need to exceed a certain threshold before they translate into measurable differences in learning.

Lack of significant advantage for the ST is in line with studies that reported no benefits for search-based tasks (e.g., Chen, 2017; Huang, 2018; Hulstijn et al., 1996; Snoder, 2017). Yang and Cao (2021), in contrast, found better recall for a search-based task in delayed retention. These studies together may indicate that the benefits of search do not rely only on dictionary consultation, but on how search was employed within a task and the extent to which it promotes meaningful engagement with lexical items.

The absence of superior vocabulary gains from the ST condition may appear, at first glance, to be inconsistent with the extensive literature on dictionary use (Cancino, 2023; Dziemianko, 2010, 2017; Laufer & Hill, 2000; Lo, 2024, 2025; Uchihara et al., 2022; Udry & Berthele, 2025). However, these studies addressed a different question. Most examined whether dictionary consultation facilitates vocabulary learning relative to no lexical support or compared the effectiveness of different dictionary formats. In contrast, the present study compared dictionary consultation with direct provision of meanings while holding other variables of the model equal. The findings, therefore, do not suggest that dictionary use is ineffective as participants in the search condition recalled about half of the target-word meanings in the immediate posttest. The findings instead suggest that searching for meanings may not provide additional benefits when learners are provided with equally accessible lexical support through glosses.

Several explanations may clarify why the search condition was comparably beneficial to the no search condition. The first explanation pertains to the implementation of dictionary consultation which may have reduced opportunities for connecting form and meaning during reading. Participants in the NST encountered target words and their meanings simultaneously within the text through marginal glosses whereas learners in the ST group consulted the dictionary, recorded meanings, and then proceeded to read the passage. This temporal and modal separation may have weakened the establishment of form–meaning connections. Similar observations could be found in Udry and Berthele (2025), who found advantages for online dictionary over paper dictionary use during on-screen reading, and in Lo (2024), who showed that isolated word lookups were less beneficial than looking up the words while situated in their contexts. Another issue that could support this point of view was that many participants in the ST group associated correct meanings with incorrect word forms.

Furthermore, some participants in the ST group relied during lookup on the dictionary Arabic translations while those in the NST group were provided only with English definitions.

This may have affected learning outcomes based on previous research that associated better vocabulary learning with consulting monolingual dictionaries (Zhang et al., 2021).

In sum, the results support the prediction of the framework for immediate vocabulary learning and offer qualified support for its prediction regarding retention. The findings suggest that the predictions of the ILH Plus may be most evident when differences in task effectiveness indices are sufficiently large.

5.3 To what extent does word length contribute to incidental vocabulary learning, and does it moderate the predictions of the ILH Plus?

The CLMM for immediate recall (Table 2) revealed a significant effect of word length. Compared with short words, the odds of successful recall were significantly lower for medium words (β = -1.097, SE = 0.291, z = -3.77, p < .001) and long words (β = -1.645, SE = 0.300, z = -5.49, p < .001). The odds ratios indicate that the odds of successfully recalling short words immediately after the task were approximately 3.0 times higher (OR = 2.99) than for medium words, and 5.2 times higher (OR = 5.18) than for long words.

As shown in Table 2, the effect of word length became even more pronounced on the delayed posttest. Compared with short words, the odds of successful recall remained significantly lower for medium words (β = -1.677, SE = 0.655, z = -2.56, p = .010) and long words (β = -2.471, SE = 0.685, z = -3.61, p < .001). The odds of retaining short words after the two-week delay were approximately 5.3 times higher (OR = 5.35) than for medium words, and 11.8 times higher (OR = 11.83) than for long words.

Further pairwise comparisons (see Table 4) showed that in the immediate posttest short words were recalled better than medium and long words, and medium words were retained better than long words. In the delayed posttest, the comparisons indicated that short words were retained better than medium and long words while there was no difference between medium and long words.

For moderation effects (see Table 3), the interaction between task condition and word length did not reach statistical significance at immediate posttests (medium words: β = -0.602, SE = 0.459, z = -1.31, p = 0.190; long words: (β = -0.490, SE = 0.473, z = -1.04, p = .301). The interaction again did not reach statistical significance at delayed posttest (medium words: (β = -0.806, SE = 0.567, z = -1.42, p = 0.155); long words: (β = -1.373, SE = 0.717, z = -1.91, p = 0.056). This shows that the impact of word length did not differ significantly depending on whether participants completed the ST or the NST.

Table 3. CLMMs Predicting Main Effects and Interaction Terms on Immediate and Delayed Recall

Predictor	Immediate Recall – β (SE)	Immediate Recall – z	Immediate Recall – p	Delayed Recall – β (SE)	Delayed Recall – z	Delayed Recall – p
Search (ST)	0.819 (0.369)	2.22	.027	0.046 (0.378)	0.12	.903
Medium word	-0.779 (0.382)	-2.04	.041	-1.251 (0.715)	-1.75	.080
Long word	-1.389 (0.395)	-3.51	< .001	-1.853 (0.742)	-2.50	.012
Simple WM	0.173 (0.177)	0.98	.328	0.012 (0.205)	0.06	.955
Complex WM	0.163 (0.182)	0.89	.372	0.312 (0.203)	1.54	.125
ST × Medium word	-0.602 (0.459)	-1.31	.190	-0.806 (0.567)	-1.42	.155
ST × Long word	-0.490 (0.473)	-1.04	.301	-1.373 (0.717)	-1.91	.056
ST × phonological WM	0.223 (0.277)	0.80	.422	0.375 (0.331)	1.13	.258
ST × executive WM	-0.122 (0.248)	-0.49	.624	-0.506 (0.292)	-1.73	.083
Random intercept						
Participant	0.436 (0.660)	0.436 (0.660)	0.436 (0.660)	0.420 (0.648)	0.420 (0.648)	0.420 (0.648)
Word	0.034 (0.184)	0.034 (0.184)	0.034 (0.184)	0.348 (0.590)	0.348 (0.590)	0.348 (0.590)

Table 4. Pairwise Comparisons Among Word-Length Categories

Comparison	Immediate	Delayed
Short vs. Medium	$p < .001$	$p < .001$
Short vs. Long	$p < .001$	$p < .001$
Medium vs. Long	$p = .028$	$p = .101$

Overall, these findings show that word length explained great variation in both vocabulary acquisition and retention independent of the predictions of the ILH Plus. Short words were consistently recalled better than longer words across the two posttests. These findings align with previous studies that reported negative effects of word length measured in letters, phonemes, or syllables on vocabulary knowledge and vocabulary learnability (Alsaif & Milton, 2012; Reynolds et al., 2015; Vanbuel, 2025; Willis & Ohashi, 2012). They are also compatible with Barclay and Pellicer-Sánchez's (2021) finding that longer words impose a greater learning burden. The findings, in fact, highlight the importance of word length as a word characteristic in incidental vocabulary acquisition and retention.

However, the results of this study run counter to other studies (Mulder et al., 2019; Urdaniz & Skoufaki, 2022) in which word length was not found a predictor of vocabulary knowledge when other word features were controlled. The results are also different from Puimège and Peters (2019), who reported a positive correlation between word length and incidental learning of single words, and from Masrai and Milton (2015) and Zeng et al. (2022), who reported weak correlations that disappeared in regression analyses. These inconsistencies may be due to differences in participant population, word knowledge assessment tools, and the extent to which other word characteristics were considered.

Beyond the overall effect of word length, the delayed posttest revealed an interesting pattern. Although short words continued to be recalled significantly better than both medium and long words, the difference between medium and long words was no longer statistically significant. One explanation of this pattern can be drawn from the results of Barclay and Pellicer-Sánchez's (2021) who found that word length predicted learning burden but not learning decay. This means that longer words required greater effort to learn initially but were not forgotten more rapidly than shorter words. This may explain why recall of long words became comparable to that of medium words in the long term.

Despite this great impact of word length on learning outcomes, it did not moderate the relative effectiveness of the two learning tasks. The advantage of shorter words was evident in both task conditions, indicating that word length affected the overall learning difficulty but not the relative effectiveness of the ST and NST. The absence of moderation is theoretically important because it suggests that variables beyond those specified in the model contribute to incidental vocabulary learning and indicate that lexical characteristics may represent an important source of variation not yet captured by the model.

These findings also make an important contribution to the broader literature on word length. While most previous studies examined word length using vocabulary-size or vocabulary-knowledge measures (e.g., Alsaif & Milton, 2012; Reynolds et al., 2015; Vanbuel, 2025), the present study examined incidental acquisition and retention of previously unfamiliar words encountered during a single learning task. Therefore, the results indicate that word length affects not only accumulated vocabulary knowledge but also the initial acquisition and retention of new lexical items.

5.4 To what extent does working memory contribute to incidental vocabulary learning, and does it moderate the predictions of the ILH Plus?

In the immediate recall in the main effects model (Table 2), phonological WM demonstrated a positive main effect that approached statistical significance ($\beta = 0.265$, SE = 0.136, $z = 1.95$, $p = .051$). This marginal trend suggests that participants with greater phonological WM capacity tended to achieve slightly higher immediate vocabulary acquisition scores, although the effect fell just short of the conventional threshold for statistical

significance ($p < .05$). In contrast, executive WM did not significantly contribute to immediate recall ($\beta = 0.094$, SE = 0.123, $z = 0.77$, $p = .444$).

By the two-week delayed posttest, neither measure of WM predicted vocabulary retention significantly (Table 2). Phonological WM was not a significant predictor of delayed recall ($\beta = 0.172$, SE = 0.163, $z = 1.06$, $p = .290$), and executive WM showed no significant effect as well ($\beta = 0.065$, SE = 0.147, $z = 0.45$, $p = .656$). These findings suggest that the marginal influence of phonological WM observed during immediate recall did not persist over time. Consequently, neither phonological storage capacity nor executive WM accounted for individual differences in long-term vocabulary retention.

As Table 3 shows, the interaction between task condition and phonological WM was not statistically significant for either the immediate posttest ($\beta = 0.223$, SE = 0.277, $z = 0.80$, $p = .422$) or the delayed posttest ($\beta = 0.375$, SE = 0.331, $z = 1.13$, $p = .258$). Similarly, the interaction between task condition and executive WM was not statistically significant on the immediate posttest ($\beta = -0.122$, SE = 0.248, $z = -0.49$, $p = .624$) or the delayed posttest ($\beta = -0.506$, SE = 0.292, $z = -1.73$, $p = .083$). These results indicate that the relationship between WM capacity and vocabulary recall did not differ significantly between the ST and NST conditions. In other words, there was no evidence that the effect of either phonological or executive WM on vocabulary acquisition or retention depended on task type.

Since all interactions were found insignificant and adding the interaction terms did not improve the models fit (immediate: AIC = 905.001, delayed: AIC = 592.283), the main effects models (immediate: AIC = 899.629, Delayed: AIC = 592.120) were used as the primary source of results.

To summarize, the findings suggest that phonological WM may play a greater role than executive WM in immediate vocabulary acquisition although neither effect reached statistical significance. For retention, neither phonological WM capacity nor executive WM accounted for differences in long-term retention. In addition, no moderation effects on the effectiveness of task conditions were found for either WM measure.

The literature supports the finding that WM predict only initial vocabulary acquisition but not retention. For example, Yang et al. (2017), Malone (2018) and Montero Perez (2020) all found WM effect on immediate learning. This suggests that for initial encoding learners may depend more heavily on the phonological WM to notice and process new lexical items. For retention, learners may rely more on the knowledge that has been consolidated in the long-term memory, and the process of consolidation may depend on other factors beyond the ones needed for initial learning.

However, the results of this study are in contrast with those from Teng (2024, 2025) and Lim and Lee (2025) who found that immediate learning and retention are significantly influenced by WM. The present study also differs from the studies by Jung et al. (2025) and Suárez et al. (2021) in which no effects of WM were observed. The differences in the results may come from methodological differences. One important difference concerns the measures of WM. Reading-span and listening-span tests were mostly used by previous studies. These are language-based tests and therefore may resemble language learning better than the digit-span tests which may have not captured WM capacity accurately (Lim & Lee, 2025).

Another methodological difference is relevant to the measurement tools of vocabulary knowledge. Recognition tests were used in many previous studies while a meaning-recall test was employed in this study. Recall tests impose significantly higher cognitive retrieval demands because learners must independently generate the target meaning without alternative choices. As a result, recall tests can constrain overall performance levels and potentially attenuate the observable contributions of WM that might emerge under less demanding recognition tests.

Although the evidence from the present study was insufficient to establish a statistically significant effect, the positive association between phonological WM and immediate recall is broadly consistent with Teng (2025). This finding can be explained in light of the cognitive

demands of the learning task which may have relied primarily on storing and maintaining information with relatively no manipulation. In both tasks, learners encountered new lexical items and had to maintain information about them while completing the reading task. This need to store lexical information may have depended more on the phonological loop which is responsible for storing verbal information temporarily.

Finally, neither phonological nor executive WM moderated the effects of task condition on vocabulary learning. Differences in WM capacity did not make one task more effective than the other on vocabulary learning. The absence of moderation effects entails that WM contributes to vocabulary learning independently from the factors specified in the ILH Plus. Finding no moderation may also suggest that the two tasks imposed the same cognitive demands leaving little opportunity for WM to affect performance differently.

6. Conclusion

The study contributes to vocabulary learning literature by being one of the first evaluations of the ILH Plus (Yanagaisawa & Webb, 2022). The study tested the model's predictive accuracy regarding the contribution of the search component and the extent to which word- and learner-related factors account for additional variance in vocabulary learning. A task with dictionary search yielded no additional benefit for immediate acquisition or delayed retention when compared to a task that provided word meanings directly. This indicates that forcing learners to look up words does not necessarily improve learning when alternative tasks match its overall cognitive load through the other components of the ILH Plus. Therefore, the findings support the relatively small weight assigned to search within the model under the learning conditions examined.

The study also shows that variation in vocabulary learning was not fully accounted for in the ILH Plus framework. Word length appeared as a robust and consistent predictor of both immediate and delayed recall of vocabulary explaining great variance in learning outcomes while learners' WM played a limited role. In addition, finding no moderation effects of word length and WM on the predictions of the model suggests that the predicted effectiveness of tasks was stable across the investigated conditions.

These insights carry both theoretical and practical implications. Theoretically, the study highlights the need for future research to investigate whether lexical characteristics can enhance the explanatory power of the ILH Plus framework in different learning contexts. The study also encourages future research to examine the relative weight of the other components of the framework and suggest refinements accordingly. From a pedagogical perspective, the findings show that increasing task demands through dictionary search alone may not be sufficient to improve incidental vocabulary learning. Therefore, teachers and materials developers may consider focusing more on word properties such as word length since it has a greater impact on vocabulary learning than the presence or absence of search.

As with any empirical investigation, the current study has several limitations that need to be acknowledged. First, all participants were female university EFL learners, and this may limit the generalizability to male learners and to younger EFL populations. Second, the study was conducted in one Saudi university with all participants sharing Arabic as their L1. Therefore, the findings may not extend to learners from other educational settings or linguistic backgrounds, where instructional practices and L1 characteristics may influence vocabulary learning. Another limitation concerns including participants only from low-intermediate proficiency level. Task involvement, word length, and WM are all factors that may affect learners with higher or lower proficiency levels differently. Future studies are, therefore, still needed to test the ILH Plus in different educational settings and with participants of different characteristics. Another limitation relates to the digit-span tasks used to measure WM capacity. Although these tasks are validated and widely used, they may not capture all dimensions of WM capacity. Future studies should therefore combine digit-span tasks with additional WM measures, such as reading-span tasks, to provide a more comprehensive assessment of WM

capacity. Lastly, convenience sampling employed in this study may limit the representativeness of the sample as participants were selected based on accessibility rather than randomly. Despite these limitations, the findings offer a foundation for future research examining how task-, word-, and learner-related factors jointly influence vocabulary learning and retention. Such work may contribute to the continued refinement of the ILH Plus framework and to the design of more effective vocabulary learning tasks.

7. References

Alsaif, A., & Milton, J. (2012). Vocabulary input from school textbooks as a potential contributor to the small vocabulary uptake gained by English as a foreign language learners in Saudi Arabia. *The Language Learning Journal*, 40, 21 - 33. https://doi.org/10.1080/09571736.2012.658221.

Baddeley, A. (2003). Working memory and language: An overview. *Journal of Communication Disorders, 36*(3), 189–208. https://doi.org/10.1016/S0021-9924(03)00019-4

Baddeley, A. (2015). Working memory in second language learning. In Z. Wen, M. Mota, & A. McNeill (Eds.), *Working memory in second language acquisition and processing* (pp. 17–28). Multilingual Matters.

Baddeley, A., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *Recent advances in learning and motivation* (pp. 47–89). Academic Press.

Barclay, S., & Pellicer-Sánchez, A. (2021). Exploring the learning burden and decay of foreign language vocabulary knowledge: The effect of part of speech and word length. ITL-International Journal of Applied Linguistics, 172(2), 259-289. https://doi.org/10.1075/itl.20011.bar

Cancino, M. (2023). Topic interest increases L2 incidental vocabulary learning and effective dictionary look-up behaviour. *Learning and Motivation*. https://doi.org/10.1016/j.lmot.2023.101920

Chen, Y. (2017). Dictionary use for collocation production and retention: A CALL-based study. *International Journal of Lexicography*, *30*(2), 225-251. https://doi.org/10.1093/ijl/ecw005

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*, 671–684. https://doi.org/10.1016/S0022-5371(72)80001-X

Dziemianko, A. (2010). Paper or electronic? The role of dictionary form in language reception, production and the retention of meaning and collocations. *International Journal of Lexicography*, *23*(3), 257-273. https://doi.org/10.1093/ijl/ecp040

Dziemianko, A. (2017). Dictionary form in decoding, encoding and retention: Further insights. *ReCALL*, *29*(3), 335–356. doi:10.1017/S0958344017000131

Eckerth, J., & Tavakoli, P. (2012). The effects of word exposure frequency and elaboration of word processing on incidental L2 vocabulary acquisition through reading. *Language Teaching Research*, *16*(2), 227-252. https://doi.org/10.1177/1362168811431377

Folse, K. S. (2006). The effect of type of written exercise on L2 vocabulary retention. *TESOL Quarterly, 40*(2), 273-293. https://doi.org/10.2307/40264523

Hazrat, M., & Read, J. (2022). Enhancing the involvement load hypothesis as a tool for classroom vocabulary research. *TESOL Quarterly, 56*(1), 387–400. https://doi.org/10.1002/tesq.3051

Huang, S. Y. (2018). A verification of involvement load hypothesis on Chinese adult English learners. *International Journal of English Linguistics, 8*(5), 125-134. https://doi.org/10.5539/ijel.v8n5p125

Hulstijn, J. H., Hollander, M., & Greidanus, T. (1996). Incidental vocabulary learning by advanced foreign-language students: the influence of marginal glosses, dictionary use, and reoccurrence of unknown words. The Modern Language Journal, 80(3), 327-339. https://doi.org/10.2307/329439

Juffs, A., & Harrington, M. (2011). Aspects of working memory in L2 learning. *Language Teaching, 44*(2), 137–166. https://doi.org/10.1017/S0261444810000509

Jung, J., Zhang, W., & Lee, M. (2025). The role of working memory and attention control in incidental L2 vocabulary learning from reading-while-listening. *ITL - International Journal of Applied Linguistics*. https://doi.org/10.1075/itl.24001.jun

Keating, G. D. (2008). Task effectiveness and word learning in a second language: The involvement load hypothesis on trial. *Language Teaching Research*, 12(3), 365-386. https://doi.org/10.1177/1362168808089922

Kim, Y. (2008). The role of task-induced involvement and learner proficiency in L2 vocabulary acquisition. *Language learning*, *58*(2), 285-325. https://doi.org/10.1111/j.1467-9922.2008.00442.x

Komuro, R. (2023). An empirical examination of incidental vocabulary learning formula: Comparison between sentence- and composition-writing tasks. *JACET Journal*, *67*, 13–31. https://doi.org/10.32234/jacetjournal.67.0_13

Laufer, B., & Hill, M. (2000). What Lexical Information Do L2 Learners Select in a CALL Dictionary and How Does It Affect Word Retention?. *Language Learning & Technology*, *3*(2), 58–76. https://doi.org/10.64152/10125/25073

Laufer, B., & Hulstijn, J. H. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics, 22*, 1–26. https://doi.org/10.1093/applin/22.1.1

Lim, S., & Lee, S. (2025). How do text type, prior vocabulary knowledge, and working memory capacity affect second language incidental vocabulary learning through reading?. *Studies in Second Language Learning and Teaching*. https://doi.org/10.14746/ssllt.33236

Linck, J. A., Osthus, P., Koeth, J. T., & Bunting, M. F. (2014). Working memory and second language comprehension and production: A meta-analysis. *Psychonomic bulletin & review*, *21*(4), 861-883. https://doi.org/10.3758/s13423-013-0565-2

Liu, S., & Reynolds, B. L. (2022). Empirical support for the involvement load hypothesis (ILH): a systematic review. *Behavioral Sciences*, *12*(10), 354. https://doi.org/10.3390/bs12100354

Lo, S. (2024). The effects of NMT as a de facto dictionary on vocabulary learning: a comparison of three look-up conditions. *Computer Assisted Language Learning*, 1-21. https://doi-org.sdl.idm.oclc.org/10.1080/09588221.2024.2310285

Lo, S. (2025). Neural machine translation in EFL classrooms: learners' vocabulary improvement, immediate vocabulary retention and delayed vocabulary retention. *Computer Assisted Language Learning*, *38*(3), 592–611. https://doi.org/10.1080/09588221.2023.2207603

Malone, J. (2018). Incidental Vocabulary Learning in SLA: Effects of Frequency, Aural Enhancement, and Working Memory. *Studies in Second Language Acquisition*, *40*(3), 651–675. https://doi.org/10.1017/S0272263117000341

Masrai, A., & Milton, J. (2015). Word Difficulty and Learning among Native Arabic Learners of EFL. *English language teaching*, *8*(6), 1-10. http://dx.doi.org/10.5539/elt.v8n6p1.

Montero Perez, M. (2020). Incidental Vocabulary Learning Through Viewing Video. *Studies in Second Language Acquisition*, 42, 749 - 773. https://doi.org/10.1017/S0272263119000706.

Mulder, E., Van De Ven, M., Segers, E., & Verhoeven, L. (2019). Context, word, and student predictors in second language vocabulary learning. *Applied Psycholinguistics*, *40*(1), 137–166. doi:10.1017/S0142716418000504

Peters, E. (2020). Factors affecting the learning of single-word items. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 125-242). Routledge

Puimège, E., & Peters, E. (2019). Learning L2 vocabulary from audiovisual input: an exploratory study into incidental learning of single words and formulaic sequences. *The Language Learning Journal*, *47*(4), 424–438. https://doi.org/10.1080/09571736.2019.1638630

Reynolds, B., Wu, W., Liu, H., Kuo, S., & Yeh, C. (2015). Towards a Model of Advanced Learners' Vocabulary Acquisition: An Investigation of L2 Vocabulary Acquisition and Retention by Taiwanese English Majors. *Applied Linguistics Review*, 6, 121 - 144. https://doi.org/10.1515/applirev-2015-0006

Snoder, P. (2017). Improving English Learners' Productive Collocation Knowledge: The Effects of Involvement Load, Spacing, and Intentionality. *TESL Canada Journal*, *34*(3), 140-164. http://dx.doi.org/10.18806/tesl.v34i3.1277

Stoet, G. (2010). PsyToolkit – A software package for programming psychological experiments using Linux. *Behavior Research Methods*, *42*(4), 1096–1104. https://doi.org/10.3758/BRM.42.4.1096

Stoet, G. (2017). PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology*, *44*(1), 24–31. https://doi.org/10.1177/0098628316677643

Suárez, M. D. M., Gilabert, R., & Moskvina, N. (2021). The Mediating Role of Vocabulary Size, Working Memory, Attention, and Inhibition in Early Vocabulary Learning Under Different TV Genres: An Exploratory Study. *TESOL Journal*, *12*(4), e637. https://doi.org/10.1002/tesj.637

Teng, M. F. (2024). Working memory and prior vocabulary knowledge in incidental vocabulary learning from listening, reading, reading-while-listening, and viewing captioned videos. *System*, 103381. https://doi.org/10.1016/j.system.2024.103381

Teng, M. F. (2025). Effectiveness of captioned videos for incidental vocabulary learning and retention: the role of working memory. *Computer Assisted Language Learning*, *38*(1–2), 206–234. https://doi.org/10.1080/09588221.2023.2173613

Uchihara, T., Eguchi, M., & Clenton, J. (2022). The contribution of guessing from context and dictionary use to receptive and productive vocabulary knowledge: A structural equation modeling approach. *Language Teaching Research*, 0(0). https://doi.org/10.1177/13621688221122138

Udry, I., & Berthele, R. (2025). A case study of online and paper bilingual dictionaries and their impact on vocabulary learning through meaning-focused reading. *International Journal of Lexicography*, *38*(3), 219-237. https://doi.org/10.1093/ijl/ecaf004

Urdaniz, R. & Skoufaki, S. (2022). Spanish L1 EFL learners' recognition knowledge of English academic vocabulary: The role of cognateness, word frequency and length. *Applied Linguistics Review*, *13*(4), 661-703. https://doi.org/10.1515/applirev-2018-0109

Vanbuel, M. (2025). Predicting vocabulary knowledge in adult L2 learners: The role of word-level variables across educational backgrounds. *Bilingualism: Language and Cognition*. https://doi.org/10.1017/s1366728924000889.

Waters, G. S., & Caplan, D. (2003). The reliability and stability of verbal working memory measures. *Behavior Research Methods, Instruments, & Computers*, *35*(4), 550-564. https://doi.org/10.3758/BF03195534

Webb, S. (2019). Incidental Vocabulary Learning. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 225-239). Routledge.

Willis, M., & Ohashi, Y. (2012). A model of L2 vocabulary learning and retention. *The Language Learning Journal*, *40*(1), 125-137. https://doi.org/10.1080/09571736.2012.658232

Yanagisawa, A. (2020). *How can we predict incidental L2 vocabulary learning? A meta-analytic examination of the involvement load hypothesis* [Doctoral dissertation, The University of Western Ontario]. Western University Open Repository. https://uwo.scholaris.ca/items/1c0e4570-c043-4115-9bcd-e8ece095cfa5

Yanagisawa, A., & Webb, S. (2021). To what extent does the involvement load hypothesis predict incidental L2 vocabulary learning? A meta-analysis. *Language Learning*, *71*(2), 487-536. https://doi.org/10.1111/lang.12444

Yanagisawa, A., & Webb, S. (2022). Involvement load hypothesis plus: Creating an improved predictive model of incidental vocabulary learning. *Studies in Second Language Acquisition*, *44*(5), 1279-1308. https://doi.org/10.1017/S0272263121000577

Yang, Y., & Cao, X. (2021). Effects of task involvement load on L2 vocabulary acquisition and their association with language aptitude. The Asia-pacific education researcher, 30, 421-430. https://doi.org/10.1007/s40299-020-00528-8

Yang, Y., Shintani, N., Li, S., & Zhang, Y. (2017). The effectiveness of post-reading word-focused activities and their associations with working memory. *System*, *70*, 38–49. https://doi:10.1016/j.system.2017.09.012

Zeng, Y., Lu, Q., Wallace, M. P., Guo, Y., Fan, C. W., & Chen, X. (2022). Understanding Sustainable Development of English Vocabulary Acquisition: Evidence From Chinese EFL Learners. *Sustainability*, *14*(11), 6532. https://doi.org/10.3390/su14116532

Zhang, S., Xu, H., & Zhang, X. (2021). The effects of dictionary use on second language vocabulary acquisition: A meta-analysis, *International Journal of Lexicography*, Volume 34, Issue 1, March 2021, Pages 1–38, https://doi.org/10.1093/ijl/ecaa010

AUTHORS' BIO

Maram S. Almohaimeed is a lecturer in the English Language Skills Department of the Common First Year at King Saud University. She holds a master's degree in Applied Linguistics and is currently pursuing a PhD in Applied Linguistics. Her research interests include instructed second language acquisition, incidental vocabulary learning, and cognitive factors influencing L2 vocabulary development.