Teaching English as a Second Language Quarterly (TESLQ)
(Formerly Journal of Teaching Language Skills)
42(3), Summer 2023, pp. 117-146
https://doi.org/10.22099/tesl.2023.48255.3220
https://dorl.net/dor/20.1001.1.20088191.2023.42.3.6.8

Online ISSN: 2717-1604
Print ISSN: 2008-8191
Research Paper

The Contribution of Working Memory and Language Proficiency to Lexical Gain: Insights from the Involvement Load Hypothesis

Hossein Kargar Behbahani * **Seyyed Ayatollah Razmjoo** **

Abstract

This study investigated the interplay of working memory capacity and language proficiency in the context of vocabulary acquisition through word-focused tasks. The involvement load hypothesis served as a theoretical framework, positing that the degree of cognitive engagement influences lexical learning outcomes. A total of 100 EFL learners participated in this study. They were divided into distinct groups based on varying levels of working memory capacity, language proficiency, and aspects of working memory (phonological short-term memory and executive working memory). The participants engaged in receptive and productive vocabulary tasks under different experimental conditions, including reading-only, reading plus blank-fill, and reading-plus production. The results revealed intricate relationships between working memory capacity, language proficiency, and vocabulary acquisition. While the efficacy of the hypothesis varied across conditions, its predictions were influenced by the nuances of individual cognitive capabilities and language competence. Learners with higher proficiency levels acquired more lexical items, both receptively and productively. Although the difference with the phonological short-term memory was not significant, executive working memory was more facilitative of lexical gain. Additionally, the role of input modality in shaping vocabulary learning outcomes was highlighted. That is, reading plus production and blank-fill tasks resulted in more lexical gain than reading-only tasks. The study contributes to the theoretical understanding of vocabulary acquisition by underscoring the complex interplay of cognitive processes and language factors. These insights hold implications for foreign language pedagogy, guiding educators in crafting more effective interventions for enhancing both receptive and productive lexical knowledge.

Keywords: Language Proficiency, Involvement Load Hypothesis, Receptive and Productive Lexical Gain, Working Memory

* Received: 02/08/2023 Accepted: 15/09/2023

* Ph.D. Candidate, Shiraz University, Iran. hossein_july1993@yahoo.com (Corresponding Author)
** Professor, Shiraz University, Iran. arazmjoo@rose.shirazu.ac.ir

How to cite this article:
Kargar Behbahani, H., & Razmjoo, S. A. (2023). The Contribution of Working Memory and Language Proficiency to Lexical Gain: Insights from the Involvement Load Hypothesis. *Teaching English as a Second Language Quarterly (Formerly Journal of Teaching Language Skills)*, *42*(3), 117-146. https://doi.org/10.22099/tesl.2023.48255.3220

Language, in its essence, is a tapestry woven from words. Vocabulary, the foundation of language, is the thread that binds this tapestry together, giving it shape and meaning (Harmer, 2011). The significance of vocabulary in the process of acquiring a new language cannot be overstated. Mastery of vocabulary empowers language learners to articulate their thoughts, engage in meaningful conversations, comprehend written texts, and express themselves eloquently (Nation, 2001). As such, vocabulary learning is a pivotal component of language acquisition.

Effective vocabulary acquisition opens opportunities for language learners (Wilkins, 1972; Yaqubi et al., 2012). It equips them with the tools necessary to navigate diverse linguistic landscapes, fostering a deeper understanding of culture, literature, and communication (Nation & Webb, 2011; Fazilatfar et al., 2011). Moreover, vocabulary proficiency is a key determinant of language proficiency as a whole (Schmitt, 2000). Whether one aims to excel in academic pursuits, communicate effectively in professional settings, or simply connect with people from different linguistic backgrounds, a robust vocabulary is an indispensable asset (Hayati & Shahriari, 2012; Thornbury, 2002).

Recognizing the paramount importance of vocabulary in language learning, educators and researchers have dedicated extensive efforts to uncover the most efficient strategies for encouraging vocabulary development in language learners. This pursuit of effective vocabulary instruction has led to the emergence of various theoretical frameworks and hypotheses aimed at shedding light on the intricate process of lexical gain. Since it is now widely understood how important vocabulary is, there has been a lot of research on how to encourage vocabulary development in language learners. The involvement Load Hypothesis (ILH) developed by Laufer and Hulstijn (2001) offered a framework for scrutinizing activities from this perspective and made an important addition to vocabulary study.

Research employing ILH has substantiated the impacts of different word-focused tasks on lexical gain (Nassaji & Hu, 2012; Teng & Zhang, 2021). These studies have revealed that tasks characterized by a greater involvement load yield more significant advancements in lexical gain. Consequently, the findings affirm that the extent of information processing applied to target words plays a decisive role in determining vocabulary acquisition outcomes. Researchers have urged educators to create tasks that increase the involvement load to improve lexical gain. According to Laufer (2003), one reason for creating these tasks is that learners are more likely to pay attention to new words when intricate processing is taking place, boosting their likelihood of picking up new words.

After analyzing the findings of earlier research in the area of incidental vocabulary learning, Laufer and Hulstijn (2001) proposed the ILH based on three assumptions. The first assumption is that three factors affect how well a task is done. One of them is a motivator called need, which is characterized as the learners' desire to comprehend language. The other two, search and evaluation, are cognitive elements. Search is the process of looking up a word's definition in a dictionary. Lastly, learners' word elaboration is called evaluation. These L2 researchers specified a range of prominence for each component. When the desire to understand a word is self-imposed, the need is generated to a strong degree (++). When understanding a word is imposed externally, it is moderate (+). Without these factors, need is not at all induced (-).

The task-induced involvement load is defined by Laufer and Hulstijn (2001) as the sum of the components with respect to their degree. This leads to the second premise: words processed with a higher involvement load are learned more effectively under similar circumstances. The third assumption thus asserts that assignments with a larger involvement load are better for acquiring new lexical items. The same engagement load should, therefore, result in equal effectiveness.

It is widely acknowledged that individual differences play a significant role in the process of acquiring a new language (Ellis, 2015). One cognitive factor that has the potential to impact the speed of language learning is working memory (WM). Baddeley (2003) has defined WM capacity as the ability to temporarily store and process information, making it a critical component for efficient real-time language processing (Miyake & Friedman, 1998). Despite the importance of this trait, certain studies have not established a clear and positive correlation between WM and language learning (Crossley & Kim, 2019; Kormos & Trebits, 2011).

The correlation between language proficiency and vocabulary learning is intrinsic and multifaceted. As language learners advance in their proficiency levels, they typically exhibit greater receptivity and capacity for acquiring new words (Zareva et al., 2005). This correlation extends to both receptive and productive vocabulary. Proficient language users tend to possess a broader lexical repertoire, enabling them to comprehend complex texts, express nuanced ideas, and engage in more meaningful interactions. Furthermore, higher language proficiency often correlates with an enhanced ability to infer word meanings from context, facilitating incidental lexical gain (Nizonkiza, 2011). However, it is essential to note that the relationship is reciprocal; vocabulary enrichment, in turn, contributes to language proficiency growth, creating a mutually reinforcing cycle in the journey of language acquisition (Tilfarlioglu & Bozgeyik, 2012).

The three factors explained above may not be the only important considerations affecting the hypothesis's predictive power. Additionally, certain arguments have been made regarding the ILH, suggesting that factors such as time spent on tasks, learners' proficiency levels, and the frequency of exposure to target words might influence the assumptions underlying the hypothesis (Hazrat & Read, 2021). Despite the growing interest in ILH, further research seems necessary to address the hypothesis's capacity to predict the future. Working memory capacity (WMC) and level of English proficiency are likely to determine vocabulary learning success. In the meantime, Wen (2016) claims that phonological short-term memory and executive WM are associated with different facets of L2 learning. These two functions of WM could potentially forecast distinct categories of vocabulary knowledge, including receptive and productive knowledge. Moreover, because lexical gain is an incremental endeavor, it is necessary to see whether the vocabulary learned as a result of ILH as a tool for classroom vocabulary research can be retained in the memory for a long time. Thus, this longitudinal study looks at how learner-related features, such as different aspects of WM and English proficiency level, may affect the results of word-focused tasks in three conditions (i.e., reading only, reading plus blank-fill, and reading plus production) in terms of vocabulary learning.

The objectives of this study are fivefold. Firstly, we study word-focused tasks effect in three conditions (i.e., reading only, reading plus blank-fill, and reading plus production) on receptive and productive lexical gain. Secondly, the study looks into the effect of WMC on receptive and productive lexical gain. Thirdly, the study will examine the effects of different aspects of WM on receptive and productive lexical gain. Fourthly, this study examines the mediating role of language proficiency on receptive and productive lexical gains. Lastly, the study investigates if the effects of WM and language proficiency on receptive and productive lexical gain are durable over time.

1. How do word-focused tasks in three conditions (i.e., reading only, reading plus blank-fill, and reading plus production) affect receptive and productive lexical gain?
2. How does working memory capacity affect receptive and productive lexical gain?
3. What aspect of working memory (i.e., phonological short-term memory and executive working memory) does affect receptive and productive lexical gain?
4. What is the effect of language proficiency on receptive and productive lexical gain?
5. Are working memory and language proficiency effects durable on receptive and productive lexical gain?

As mentioned above, few studies have investigated the effects of WMC and language proficiency on vocabulary knowledge within the ILH framework. Thus, this study seems to be an innovation because it deals with different aspects of WMC and how language proficiency affects receptive and productive lexical gain in the long run. It is also hoped that the knowledge that will be gained as a result of this current investigation adds to the literature, fills the gap, and provides insights for both materials developers and foreign language teachers on how receptive and productive vocabulary aspects can better be acquired.

Literature Review

Both ILH and technique feature analysis are vocabulary learning hypotheses aiming to give prominence to concepts like awareness, motivation, and engagement. To evaluate the predictive capabilities of these in lexical gain, Jafari Gohar et al. (2018) divided a cohort of 90 highly proficient EFL pupils into three tasks: sentence construction, composition, and comprehension. It was discovered that technique feature analysis was a better predictor of the change in score from the pre-test to the post-test than ILH was of the during-task activity.

In another research inquiry, Namaziandost et al. (2020) compared the effects of high involvement load and low involvement load on Iranian sophomore EFL learners' vocabulary development. Their experiment involved two intact sophomore BA courses. It employed a cross-sectional design comprising both comparison and treatment groups. The first reading comprehension test had to be taken by the first experimental group, which had a high engagement burden. The first Vocabulary Knowledge Scale (VKS) was used to determine if any gain had taken place in the targeted items selected from the reading. The second experimental group (the one with a low engagement burden) received the second reading comprehension task at the same time as the first experimental group but in a different location. They were given the second VKS after two weeks. It was discovered after data analysis that exposing learners to high levels of participation load can significantly contribute to the development of English vocabulary. Additionally, it was revealed that although participants remembered the vocabulary learned with high levels of involvement load better than those with low levels of involvement load, there was no significant difference in retention. These L2 researchers concluded that vocabulary trainers and language teachers can use these findings to create efficient reading exercises with the right degree of efficiency.

Another study that investigated the efficacy of ILH on lexical gain is that of Teng and Zhang (2021). This study looked at how engagement load-based activities affect learning vocabulary in a foreign language and how much task effects may be predicted by learners' metacognition. A total of 120 Chinese ESL learners were randomly assigned to one of four distinct conditions: reading, reading plus gap-fill, reading plus writing, and reading plus writing with the aid of a digital dictionary. The VKS was modified to assess the effects of the conditions. The knowledge and regulation of learners' metacognition were evaluated using the Metacognitive Awareness Inventory. Reading + writing, reading + gap-fill, and reading-only groups all exhibited suboptimal performance in terms of receptive and productive lexical gain. Conversely, using a digital dictionary helped learners achieve the highest level of performance.

The investigation of the ILH efficacy is not limited to the above-mentioned researchers. For instance, Kaivanpanah et al. (2020) purported to determine how input-based and output-based tasks with various and equivalent degrees of engagement loads affected incidental vocabulary learning in Iranian EFL learners. One hundred twenty pre-intermediate EFL students from five Iranian English language schools participated in the course. In Phase 1, participants completed tasks that were equally involved in both input and output. They were given both input- and output-based activities in phase 2, although the input-based tasks' participation burden was larger. Finally, output-based activities with increased participation loads were given to the participants in phase 3. Following the completion of the tasks and one week after the post-tests, respectively, immediate post-tests and delayed post-tests were given to determine the level of vocabulary learning. The results revealed that vocabulary acquisition and retention, both in the post-test and the delayed post-test, were enhanced when learners engaged in assignments that balanced input and output demands. Furthermore, there was no significant impact on vocabulary learning and retention in the post-test and delayed post-test when input-based tasks with increased involvement loads were employed. Finally, it was observed that output-based tasks with higher involvement loads significantly positively influenced learners' lexical gains in the long run.

In an interesting study, Kivrak and Gokmen (2019) looked into how tasks impact vocabulary learning at various skill levels with the same engagement load but distinct input modalities (written vs. audiovisual). Two sources of input and three involvement loads were used in six vocabulary problems that were completed by 236 Turkish (lower- or upper-intermediate) EFL students. Two tasks required only reading or video comprehension, while the other four required gap-filling or the creation of sentences using eight target words. The target form or meaning was necessary for vocabulary post-tests that measured both productive and receptive word knowledge. The percentage/number of the proper forms and meanings was counted to obtain the data. The findings showed that regardless of input type, sentence writing was more effective than gap filling for receptive word knowledge at both levels, but that more knowledge was fostered among upper-intermediate students. For both levels, the combination of gap-filling with written input and sentence composition with audiovisual input was more productive in terms of word knowledge. While these results underscore the importance of input modality in productive word knowledge, they also partially validate the expectations for receptive word knowledge.

Ansarian and Kazemipour Khabbazi (2021) delved into the realms of WM and its relationship with three tasks of varying loads. Additionally, they examined the comparative impacts of single- and dual-annotation modes on the passive and active lexical development of proficient EFL learners. The study involved 204 participants who were exposed to annotations for 20 lexical items while listening to an expository passage. The researchers employed Paribakht and Wesche's (1996) VKS as the pre-and post-test, and the learners' WM abilities were assessed using a listening span task. The findings suggested that dual-mode annotations are more effective than single-mode annotations for both passive and active vocabulary learning. There was no relationship between WM and input method because the participants who had stronger WM capacity considerably outperformed the others in terms of active vocabulary learning for both textually and visually annotated items. The vocabulary tests' highest scoring task was writing sentences; however, there was no discernible difference between the cloze deletion and paragraph writing tasks. Therefore, the ILH was only partially supported by the results. The current study supports earlier research that found a modality effect on vocabulary learning and demonstrates that task type and WM are important variables in EFL vocabulary acquisition.

Perhaps the latest article on the ILH is that of Teng (2022), published in the journal of *RELC*. This researcher examined how three-word-focused workout conditions affect the acquisition of new vocabulary. The involvement load concept served as the foundation for the workout development. Additionally, this study investigates how individual characteristics (such as WMC and second-language English competence level) impact vocabulary acquisition outcomes. The three exercise conditions (reading comprehension plus marginal glosses, reading plus gap-fill, and reading plus sentence writing) were distributed evenly and randomly to 180 Chinese students. To assess vocabulary increases before and after tests, the VKS was modified. To evaluate learners' WM, an n-back task was created. The results showed that the sentence-writing group, followed by the gap-fill group and then the reading comprehension group, produced the best results in vocabulary learning. The results of a general linear model showed that learners' WM and English ability levels significantly influenced their vocabulary growth. This difference between Teng's study with ours is that we have focused on different aspects of WM on receptive and productive lexical gain. However, he has just focused on WMC's impact on vocabulary learning. Additionally, unlike him, we have focused on different tasks affecting the ILH framework on vocabulary learning in the long run.

To sum up, evidence regarding the efficacy of the ILH as a vocabulary learning tool is mixed with some L2 researchers finding support for its effectiveness (e.g., Namaziandost et al., 2020; Teng and Zhang, 2021; Teng 2022, etc.) and some L2 researchers (e.g. Ansarian and Kazemipour Khabbazi, 2021; Jafari Gohar et al., 2018) showing that the ILH is not an effective tool for lexical growth. Additionally, sorting through the literature reveals that input modality affects the way the ILH predicts vocabulary growth (for example, see Ansarian and Kazemipour Khabbazi, 2021; Kivrak and Gokmen, 2019). Additionally, although Teng (2022) studied the effect of language proficiency and WMC, he acknowledged that different aspects of WM (e.g., phonological loop and central executive), which have not been studied in his research, can affect how ILH predicts lexical gain. Thus, all these issues show that there is still some room for further research, and this study seeks to examine several components of WM and how linguistic competence influences receptive and productive lexical acquisition. Additionally, it is hoped that the information gleaned from this investigation will fill a knowledge gap, add to the body of knowledge, and offer guidance to materials designers and foreign language instructors on how receptive and productive vocabulary aspects will be more effectively learned.

Method

The study is a quasi-experimental quantitative research and uses a nonrandomized pre-test, treatment, post-test, and delayed post-test design. The vocabulary pre-test and WM test were administered in weeks one and two, respectively. The treatment was followed by an immediate post-test on vocabulary in week four and a delayed post-test in week six.

Setting and participants

Four preexisting EFL classes from Shahid Motahari Public Senior High School in Behbahan, Iran, were chosen for the data collection procedure through a convenient sampling procedure. All the participants were Iranian, with Persian as their native language. It should be mentioned that every participant was a man. There were three experimental groups and a comparison group. The participants ranged in age from 16 to 18. There were 25 participants in each group.

Instruments and materials

In the current study, the Vision 2 coursebook published by Iran's Ministry of Education was selected as the instructional material. Then, participants' language proficiency was measured using an Oxford Quick Placement Test (OQPT). A pre-test (i.e., a teacher-made test containing the would-be-taught words) was given to the subjects to ensure that nobody knew the words to be taught beforehand. At this stage, two tests of WM were given. Firstly, a reading span test developed and validated by Shahnazari (2013) allowed the researcher to determine the participants' WMC and their complex executive WM. Secondly, a non-word repetition test was also administered to determine the participants' phonological short-term memory. Then, three immediate post-tests—new tests created by the teacher—were used as the experiment developed. The first immediate post-test was a multiple-choice test. The second was a blank-fill test. The last one was a test with open-ended questions that taped into participants' production competence of the taught words. Three delayed post-tests—tests created by the instructor—were given after a two-week period to determine whether the impact of the training was long-lasting. The tests created by the researchers all followed the same format, although they differed slightly from one another. It should be noted that to ensure the construct validity of the tests, through known-group techniques, the data collector administered the tests to a group of advanced learners of English. Advanced learners' performance turned out to be different from our participants before embarking on the treatment ($p < 0.05$), hence the tests' construct validity. The data collector also asked two experts in language testing to verify the content as well as the face validity of the tests. During the construct validation phase of the study, the reliability of the instruments was also investigated through Cronbach's Alpha formula using SPSS software, which turned out to be reliable ($r = .813$).

Data collection procedures

The data collection procedures lasted six weeks. In the first two weeks, an OQPT and two tests of WM were carried out. The first test of WM was a reading span test in which students are presented with a set of sentences whose last words they need to keep in memory. The number of words participants can keep in memory shows their WM span. The non-word test measures the participants' phonological short-term memory. During this task, they were presented with non-words, and the number of the non-words that they could keep in mind shows their phonological short-term memory. In the third week, using the coursebook, new words based on the New Words and Expressions section of the third lesson of Vision 2 were introduced to the participants. It should also be noted that pictorial cues are also presented in the book so that students can learn the meaning of words more easily. The first experimental group only read the words, the second experimental group, in addition to exposure to the words, had to do some blank-filling exercises, and the last experimental group had to do some production tasks in addition to exposure to the words. The first reading task, according to Laufer and Hulstijn (2001), results in neither Need (since it is irrelevant to the task) nor Search for the meaning (because pictorial cues are provided), nor it results in any evaluation; in other words, the involvement load for the task is 0 (-+-+- = 0). As specified by Laufer and Hulstijn, the three task components are not present here, as indicated by the minus symbol (2001). The nature of this task minimizes the likelihood of incidental acquisition of target lexical items because the task's load is at its lowest level possible. For the second task, the same reading text needs to be read by the participants but without the target words. The target words appear in a random arrangement at the top of the page. The need component in this task is moderate since it was induced externally or by the activity itself. Since participants receive glossaries and do not need to look up the words, there is no search component. The researchers evaluated the candidate words against each other to assess their contextual suitability for filling in the blanks with the correct terms. The task's involvement index is 2 (+ (1) need, — (0) search, + (1) evaluation) based on the ILH. For the last task, the same vocabulary set was introduced to learners but with one significant difference. The participants had to produce sentences using the taught words in this condition. Thus, this condition resulted in need (as participants have to know the words to produce sentences), search (as the students had to search for the correct word in the mental syllabus to produce appropriate sentences), and evaluation (as their sentences were evaluated both by themselves and course instructor). This condition triggers a high involvement load. It goes without saying that the control group was not be exposed to the treatment and tasks. In the next week, a post-test was administered, and after a two-week interval, delayed post-tests were carried out to check the durability of the treatment.

Data analysis procedures

SPSS software was used to execute the statistical tests. For the first objective, which deals with the effect of word-focused tasks on receptive and productive knowledge of vocabulary, a one-way MANOVA was carried out. For the second objective dealing with the WMC effect on receptive and productive lexical gain, another one-way MANOVA was run. For the third research question that investigates the effect of different aspects of WM on receptive and productive knowledge of vocabulary, a further one-way MANOVA was conducted. For the fourth objective, examining the impact of language proficiency on receptive and productive lexical gain, another one-way MANOVA was run. Lastly, for the last research question dealing with the effect of WM and language proficiency on the receptive and productive lexical gain in the long run, a two-way MANOVA was run.

Results

We needed to run a statistical test of significance to measure the effect of word-focused tasks in three conditions on receptive and productive lexical gain. According to

Pallant (2020), when an independent variable's effect on more than one dependent variable is investigated, the conduction of MANOVA is warranted. However, before doing this, we needed to ensure the normality assumption. Thus, a one-sample Kolmogorov-Smirnov (K-S) test was run, whose results confirmed the normality assumption ($p > 0.05$).

Table 1.
Descriptive Statistics

	Group	Mean	Std. Deviation	N
Receptive Vocab Pre-test	Experimental1	3.720	1.400	25
	Experimental2	4.120	1.641	25
	Experimental 3	3.720	1.620	25
	Control	3.725	1.541	25
	Total	3.820	1.539	100
Receptive Vocab Post-test	Experimental1	6.040	1.989	25
	Experimental2	8.920	3.402	25
	Experimental 3	14.040	2.964	25
	Control	3.720	1.541	25
	Total	8.180	4.632	100
Receptive Vocab Delayed Post-test	Experimental1	5.680	1.886	25
	Experimental2	8.440	3.001	25
	Experimental 3	13.520	2.468	25
	Control	3.640	1.496	25
	Total	7.820	4.351	100
Productive Vocab Pre-test	Experimental1	3.280	1.429	25
	Experimental2	3.840	1.818	25
	Experimental 3	3.286	1.458	25
	Control	3.400	1.581	25
	Total	3.450	1.572	100
Productive Vocab Posttest	Experimental1	5.200	1.707	25
	Experimental2	8.120	3.358	25
	Experimental 3	13.360	3.264	25
	Control	3.560	1.502	25
	Total	7.560	4.537	100
Productive Vocab Delayed Post-test	Experimental1	4.960	1.619	25
	Experimental2	7.840	3.091	25
	Experimental 3	12.680	2.897	25
	Control	3.320	1.345	25
	Total	7.200	4.264	100

Table 1 presents the descriptive statistics. It shows that on the receptive pre-test of vocabulary, all groups performed approximately the same (Experimental 1 group Mean = 3.720, Experimental 2 group Mean = 4.120, Experimental 3 group Mean = 3.720, Control group mean = 3.725). It shows that all groups had the same knowledge of lexical items at baseline. The table also reveals that on the pre-test of productive vocabulary knowledge, all the groups also had the same knowledge (Experimental 1 group Mean =

3.280, Experimental 2 group Mean = 3.840, Experimental 3 group Mean = 3.286, Control group mean = 3.400). The table also presents the amount of knowledge gained in response to ILH on the immediate post-test. Accordingly, on the receptive post-test, group 3 outperformed other conditions (Experimental 1 group Mean = 6.040, Experimental 2 group Mean = 8.920, Experimental 3 group Mean = 14.040, Control group mean = 3.720). The same was the case on the productive vocabulary knowledge on time 2 (Experimental 1 group Mean = 5.200, Experimental 2 group Mean = 8.120, Experimental 3 group Mean = 13.360, Control group mean = 3.560). On the delayed post-test of receptive knowledge, performances slightly dropped; however, the third experimental condition still outperformed other groups (Experimental 1 group Mean = 5.680, Experimental 2 group Mean = 8.440, Experimental 3 group Mean = 13.520, Control group mean = 3.640). The same was also the case on time 3 on the delayed post-test of productive lexical knowledge (Experimental 1 group Mean = 4.960, Experimental 2 group Mean = 7.840, Experimental 3 group Mean = 12.680, Control group mean = 3.320). However, the fault with descriptive statistics is that it does not show whether the difference between groups is significant statistically. To this end, pairwise comparisons need to be conducted.

Table 2.
Pairwise Comparisons of different tasks effect on receptive and productive lexical gain

Dependent Variable	(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
						Lower Bound	Upper Bound
Receptive Vocab Pre-test	Experimental1	Experimental2	-.400	.439	1.000	-1.584	.784
		Experimental 3	-1.258E-017	.439	1.000	-1.184	1.184
		Control	-1.258E-017	.439	1.000	-1.184	1.184
	Experimental2	Experimental1	.400	.439	1.000	-.784	1.584
		Experimental 3	.400	.439	1.000	-.784	1.584
		Control	.400	.439	1.000	-.784	1.584
	Experimental 3	Experimental1	1.258E-017	.439	1.000	-1.184	1.184
		Experimental2	-.400	.439	1.000	-1.584	.784
		Control	.000	.439	1.000	-1.184	1.184
	Control	Experimental1	1.258E-017	.439	1.000	-1.184	1.184
		Experimental2	-.400	.439	1.000	-1.584	.784
		Experimental 3	.000	.439	1.000	-1.184	1.184
Receptive Vocab Post-test	Experimental1	Experimental2	-2.880*	.731	.001	-4.849	-.911
		Experimental 3	-8.000*	.731	.000	-9.969	-6.031
		Control	2.320*	.731	.012	.351	4.289
	Experimental2	Experimental1	2.880*	.731	.001	.911	4.849
		Experimental 3	-5.120*	.731	.000	-7.089	-3.151
		Control	5.200*	.731	.000	3.231	7.169

Dependent Variable	(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
						Lower Bound	Upper Bound
	Experimental 3	Experimental1	8.000*	.731	.000	6.031	9.969
		Experimental2	5.120*	.731	.000	3.151	7.089
		Control	10.320*	.731	.000	8.351	12.289
	Control	Experimental1	-2.320*	.731	.012	-4.289	-.351
		Experimental2	-5.200*	.731	.000	-7.169	-3.231
		Experimental 3	-10.320*	.731	.000	-12.289	-8.351
Receptive Vocab Delayed Post-test	Experimental1	Experimental2	-2.760*	.647	.000	-4.502	-1.018
		Experimental 3	-7.840*	.647	.000	-9.582	-6.098
		Control	2.040*	.647	.013	.298	3.782
	Experimental2	Experimental1	2.760*	.647	.000	1.018	4.502
		Experimental 3	-5.080*	.647	.000	-6.822	-3.338
		Control	4.800*	.647	.000	3.058	6.542
	Experimental 3	Experimental1	7.840*	.647	.000	6.098	9.582
		Experimental2	5.080*	.647	.000	3.338	6.822
		Control	9.880*	.647	.000	8.138	11.622
	Control	Experimental1	-2.040*	.647	.013	-3.782	-.298
		Experimental2	-4.800*	.647	.000	-6.542	-3.058
		Experimental 3	-9.880*	.647	.000	-11.622	-8.138
Productive Vocab Pre-test	Experimental1	Experimental2	-.560	.447	1.000	-1.763	.643
		Experimental 3	.000	.447	1.000	-1.203	1.203
		Control	-.120	.447	1.000	-1.323	1.083
	Experimental2	Experimental1	.560	.447	1.000	-.643	1.763
		Experimental 3	.560	.447	1.000	-.643	1.763
		Control	.440	.447	1.000	-.763	1.643
	Experimental 3	Experimental1	.000	.447	1.000	-1.203	1.203
		Experimental2	-.560	.447	1.000	-1.763	.643
		Control	-.120	.447	1.000	-1.323	1.083
	Control	Experimental1	.120	.447	1.000	-1.083	1.323
		Experimental2	-.440	.447	1.000	-1.643	.763
		Experimental 3	.120	.447	1.000	-1.083	1.323
Productive Vocab Posttest	Experimental1	Experimental2	-2.920*	.736	.001	-4.904	-.936
		Experimental 3	-8.160*	.736	.000	-10.144	-6.176
		Control	1.640	.736	.170	-.344	3.624
	Experimental2	Experimental1	2.920*	.736	.001	.936	4.904
		Experimental 3	-5.240*	.736	.000	-7.224	-3.256
		Control	4.560*	.736	.000	2.576	6.544
	Experimental 3	Experimental1	8.160*	.736	.000	6.176	10.144
		Experimental2	5.240*	.736	.000	3.256	7.224
		Control	9.800*	.736	.000	7.816	11.784
	Control	Experimental1	-1.640	.736	.170	-3.624	.344
		Experimental2	-4.560*	.736	.000	-6.544	-2.576
		Experimental 3	-9.800*	.736	.000	-11.784	-7.816
Productive Vocab Delayed Post-test	Experimental1	Experimental2	-2.880*	.669	.000	-4.683	-1.077
		Experimental 3	-7.720*	.669	.000	-9.523	-5.917
		Control	1.640	.669	.096	-.163	3.443

Dependent Variable	(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
						Lower Bound	Upper Bound
	Experimental2	Experimental1	2.880*	.669	.000	1.077	4.683
		Experimental 3	-4.840*	.669	.000	-6.643	-3.037
		Control	4.520*	.669	.000	2.717	6.323
	Experimental 3	Experimental1	7.720*	.669	.000	5.917	9.523
		Experimental2	4.840*	.669	.000	3.037	6.643
		Control	9.360*	.669	.000	7.557	11.163
	Control	Experimental1	-1.640	.669	.096	-3.443	.163
		Experimental2	-4.520*	.669	.000	-6.323	-2.717
		Experimental 3	-9.360*	.669	.000	-11.163	-7.557

According to the Bonferroni adjustment test, the condition difference was insignificant on the receptive pre-test ($p > 0.05$). The same was the case on the pre-test of productive vocabulary knowledge ($p > 0.05$). However, on the post-test of receptive vocabulary knowledge, the difference between experimental group 1 with the second experimental group was significant (Mean difference = - 2.880, $p < 0.05$). The difference between this group with the third experimental group was also significant (Mean difference = - 8.000, $p < 0.05$). The difference with the control group was also significant (Mean difference = 2.320, $p < 0.05$). The difference between the second experimental group and the third one was also significant (Mean difference = -5.120, $p < 0.05$). On the receptive post-test of vocabulary knowledge, other experimental groups also outstripped the control group ($p < 0.05$). On the post-test of productive lexical knowledge, the difference between experimental group 1 and group 2 was significant (Mean difference = - 2.920, $p < 0.05$). The difference between this group with the third control group was also significant (Mean difference = - 8.160, $p < 0.05$). The difference between this condition and the control group was not significant (Mean difference = 1.640, $p > 0.05$). However, the difference between other experimental conditions and the control group was significant ($p < 0.05$). At this time, the difference between the second experimental group and the third one was also significant (Mean difference = -5.240, $p < 0.05$). On the delayed post-test of receptive knowledge of vocabulary, the difference between the first experimental group and the second was significant (Mean difference = -2.780, $p < 0.05$). The disparity was also significant between the first and the second experimental group (Mean difference = -7.840, $p < 0.05$). The difference was also significant between this group and the control group (Mean difference = -2.040, $p < 0.05$). At this time, the difference between the second and the third experimental group was also significant (Mean difference = -5.080, $p < 0.05$). Lastly, on the delayed post-test of productive vocabulary knowledge, the difference was significant between the first experimental group and the second (Mean difference = -2.880, $p < 0.05$), the third experimental group (Mean difference = -7.720, $p < 0.05$), and the control group (Mean difference = 1.640, *p*

< 0.05). The second and the third experimental conditions difference was also significant (Mean difference = -4.840, $p < 0.05$).

It was essential to conduct a statistical significance test to assess WMC's impact on the development of receptive and productive vocabulary. Following the guidance of Pallant (2020), when examining an independent variable's influence on multiple dependent variables, it is appropriate to perform a MANOVA. However, before proceeding with this analysis, it was necessary to confirm that the data met the normality assumption. Consequently, a K-S test was executed, and the results affirmed that the assumption was met ($p > 0.05$).

Table 3.
Descriptive Statistics

	WM	Mean	Std. Deviation	N
Receptive Vocab Pre-test	High	4.000	1.496	51
	Low	3.632	1.577	49
	Total	3.820	1.539	100
Receptive Vocab Post-test	High	9.061	3.917	51
	Low	7.333	5.169	49
	Total	8.180	4.632	100
Receptive Vocab Delayed Post-test	High	8.734	3.786	51
	Low	6.941	4.738	49
	Total	7.820	4.351	100
Productive Vocab Pre-test	High	3.568	1.590	51
	Low	3.326	1.559	49
	Total	3.450	1.572	100
Productive Vocab Posttest	High	8.408	3.610	51
	Low	6.745	5.239	49
	Total	7.560	4.537	100
Productive Vocab Delayed Post-test	High	8.040	3.364	51
	Low	6.392	4.928	49
	Total	7.200	4.264	100

The table presents descriptive statistics of the WMC effect on receptive and lexical gain. According to the table, on the pre-test of receptive knowledge, irrespective of WMC, all participants' scores were almost the same (High-WM learners' mean = 4.000, Low-WM learners' mean = 3.632). The same was also the case at time 1 for productive knowledge (High-WM learners' Mean = 3.568, Low-WM learners' mean = 3.326). On the post-test of receptive knowledge, however, high-WM learners outperformed low-WM ones (High-WM learners' Mean = 9.061, Low-WM learners' mean = 7.333). On the post-test of productive knowledge, similarly, high-WM learners outperformed low-WM ones (High-WM learners' mean = 8.408, Low-WM learners' mean = 6.745). Furthermore, on the delayed post-test of receptive vocabulary knowledge, high-WM learners performed better than low-WM participants (High-WM learners' mean = 8.734, Low-WM learners' mean = 6.941). Similarly, on the delayed post-test of productive vocabulary knowledge, learners with higher WMC outperformed those with lower WMC (High-WM learners' Mean = 8.040, Low-WM learners' mean = 6.392). However, the fault with descriptive statistics is that it does not show whether the difference between groups is significant statistically. To this end, pairwise comparisons need to be conducted.

Table 4.
Pairwise Comparisons of working memory effect of receptive and productive lexical gain

Dependent Variable	(I) WM	(J) WM	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
						Lower Bound	Upper Bound
Receptive Vocab Pre-test	High	Low	.367	.307	.235	-.243	.977
	Low	High	-.367	.307	.235	-.977	.243
Receptive Vocab Post-test	High	Low	1.728	.915	.062	-3.544	.088
	Low	High	1.728	.915	.062	-.088	3.544
Receptive Vocab Delayed Post-test	High	Low	1.794*	.856	.039	-3.492	-.095
	Low	High	1.794*	.856	.039	.095	3.492
Productive Vocab Pre-test	High	Low	.242	.315	.444	-.383	.868
	Low	High	-.242	.315	.444	-.868	.383
Productive Vocab Posttest	High	Low	1.663	.897	.067	-3.443	.117
	Low	High	1.663	.897	.067	-.117	3.443
Productive Vocab Delayed Post-test	High	Low	1.649	.841	.053	-3.318	.020
	Low	High	1.649	.841	.053	-.020	3.318

Table 4 indicates that high- and low-WM learners difference on receptive and productive vocabulary knowledge at time 1 was not significant ($p > 0.05$). However, on the immediate receptive post-test of vocabulary knowledge, high-WM learners outperformed low-WM ones, and the difference was almost significant (Mean difference = 1.728, p = .062). Similarly, on the immediate productive post-test of vocabulary knowledge, high-WM learners outperformed low-WM ones, and the difference was almost significant (Mean difference = 1.663, p = .067). On the delayed receptive post-test of vocabulary, high-WM learners outperformed low-WM learners (Mean difference = 1.794, p = 0.039). On the delayed productive post-test of vocabulary, high-WM learners outperformed low-WM ones as well (Mean difference = 1.649, p = 0.053).

We needed to run a statistical test of significance to study how different aspects of WM affect receptive and productive lexical gain. According to Pallant (2020), when an independent variable's effect on more than one dependent variable is investigated, the conduction of MANOVA is warranted. However, before doing this, we needed to ensure the normality assumption. Thus, a K-S test was run whose results substantiated the assumption of normality ($p > 0.05$).

Table 5.
Descriptive Statistics

	WM	Mean	Std. Deviation	N
Receptive Vocab Pre-test	Executive-High	3.869	1.486	23

	Executive-Low	3.653	1.623	26
	Phonological-High	4.107	1.523	28
	Phonological-Low	3.608	1.559	23
	Total	3.820	1.539	100
Receptive Vocab Post-test	Executive -Low	8.000	4.461	23
	Executive-High	9.423	5.285	26
	Phonological-Low	6.785	3.392	28
	Phonological-High	8.652	5.122	23
	Total	8.180	4.632	100
Receptive Vocab Delayed Post-test	Executive-Low	7.478	4.088	23
	Executive-High	9.000	4.749	26
	Phonological-Low	6.500	3.532	28
	Phonological-High	8.434	4.813	23
	Total	7.820	4.351	100
Productive Vocab Pre-test	Executive-High	3.608	1.559	23
	Executive-Low	3.384	1.651	26
	Phonological-High	3.535	1.643	28
	Phonological-Low	3.260	1.483	23
	Total	3.450	1.572	100
Productive Vocab Posttest	Executive-Low	6.869	4.037	23
	Executive-High	8.500	5.375	26
	Phonological-Low	6.642	3.291	28
	Phonological-High	8.304	5.199	23
	Total	7.560	4.537	100
Productive Vocab Delayed Post-test	Executive-Low	6.347	3.773	23
	Executive-High	8.192	5.184	26
	Phonological-Low	6.428	3.060	28
	Phonological-High	7.869	4.731	23
	Total	7.200	4.264	100

Table 5 presents the descriptive statistics. It shows that on the receptive pre-test of vocabulary, all groups performed approximately the same (executive-high mean = 3.869, executive-low mean = 3.653, phonological-high Mean = 4.107, phonological-low mean = 3.608). It shows that all groups had the same knowledge of lexical items at baseline. The table also reveals that on the pre-test of productive vocabulary knowledge, all the groups also had the same knowledge (executive-high Mean =3.535, executive-low mean = 3.260, phonological-high mean = 3.535, phonological-low mean = 3.260). The table also presents the amount of knowledge gained in response to ILH on the immediate post-test. Accordingly, on the receptive post-test, executive-high WM learners outperformed other learners (executive-high mean = 9.423, executive-low mean = 8.000, phonological-high Mean = 8.652, phonological-low mean = 6.785). The same was the case on the productive vocabulary knowledge on time 2 (executive-high mean = 8.500, executive-low Mean = 6.869, phonological-high Mean = 8.304, phonological-low mean = 6.642). Similarly, on the delayed post-test of receptive knowledge, executive-high WM learners outperformed other learners (executive-high mean = 9.000, executive-low mean = 7.478, phonological-high Mean = 8.434, phonological-low mean = 6.500). The same was also the case on time 3 on the delayed post-test of productive lexical knowledge (executive-high mean = 8.192, executive-low mean = 6.347, phonological-high mean = 7.689, phonological-low mean = 6.428). However, the fault with descriptive statistics is that it does not show whether the difference between groups is significant statistically. To this end, pairwise comparisons need to be conducted.

Table 6.
Pairwise Comparisons of aspects of working memory effects on receptive and productive lexical gain

Dependent Variable	**(I) WM**	**(J) WM**	**Mean Difference (I-J)**	**Std. Error**	**Sig.[a]**	**95% Confidence Interval for Difference[a]**	
						Lower Bound	**Upper Bound**
Receptive Vocab Pre-test	Executive-High	Executive-Low	.216	.444	1.000	-.980	1.411
		Phonological-High	-.238	.436	1.000	-1.413	.938
		Phonological-Low	.261	.457	1.000	-.971	1.492
	Executive-Low	Executive-High	-.216	.444	1.000	-1.411	.980
		Phonological-High	-.453	.422	1.000	-1.591	.684
		Phonological-Low	.045	.444	1.000	-1.150	1.240
	Phonological-High	Executive-High	.238	.436	1.000	-.938	1.413
		Executive-Low	.453	.422	1.000	-.684	1.591
		Phonological-Low	.498	.436	1.000	-.677	1.674
	Phonological-Low	Executive-High	-.261	.457	1.000	-1.492	.971
		Executive-Low	-.045	.444	1.000	-1.240	1.150
		Phonological-High	.498	.436	1.000	-1.674	.677
Receptive Vocab Post-test	Executive-High	Executive-Low	1.423	1.314	1.000	-4.964	2.118
		Phonological-High	1.214	1.292	1.000	-2.267	4.696
		Phonological-Low	.652	1.354	1.000	-4.300	2.996
	Executive-Low	Executive-High	-1.423	1.314	1.000	-2.118	4.964

Dependent Variable	(I) WM	(J) WM	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a] Lower Bound	Upper Bound
		Phonological-High	-2.637	1.251	.225	-.732	6.007
		Phonological-Low	.771	1.314	1.000	-2.770	4.312
	Phonological-High	Executive-High	-1.214	1.292	1.000	-4.696	2.267
		Executive-Low	-2.637	1.251	.225	-6.007	.732
		Phonological-Low	-1.866	1.292	.911	-5.348	1.615
	Phonological-Low	Executive-High	-.652	1.354	1.000	-2.996	4.300
		Executive-Low	-.771	1.314	1.000	-4.312	2.770
		Phonological-High	-1.866	1.292	.911	-1.615	5.348
Receptive Vocab Delayed Post-test	Executive-High	Executive-Low	-1.522	1.232	1.000	-4.841	1.797
		Phonological-High	.978	1.211	1.000	-2.285	4.241
		Phonological-Low	.957	1.269	1.000	-4.376	2.463
	Executive-Low	Executive-High	-1.522	1.232	1.000	-1.797	4.841
		Phonological-High	-2.500	1.172	.213	-.658	5.658
		Phonological-Low	.565	1.232	1.000	-2.754	3.884
	Phonological-High	Executive-High	-.978	1.211	1.000	-4.241	2.285
		Executive-Low	-2.500	1.172	.213	-5.658	.658
		Phonological-Low	-1.935	1.211	.681	-5.198	1.328
	Phonological-Low	Executive-High	-.957	1.269	1.000	-2.463	4.376
		Executive-Low	-.565	1.232	1.000	-3.884	2.754
		Phonological-High	-1.935	1.211	.681	-1.328	5.198
Productive Vocab Pre-test	Executive-High	Executive-Low	.224	.455	1.000	-1.003	1.451
		Phonological-High	.073	.448	1.000	-1.133	1.279

Dependent Variable	(I) WM	(J) WM	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a]	
						Lower Bound	Upper Bound
		Phonological-Low	.348	.469	1.000	-.916	1.612
	Executive-Low	Executive-High	-.224	.455	1.000	-1.451	1.003
		Phonological-High	-.151	.433	1.000	-1.319	1.016
		Phonological-Low	.124	.455	1.000	-1.103	1.351
	Phonological-High	Executive-High	-.073	.448	1.000	-1.279	1.133
		Executive-Low	.151	.433	1.000	-1.016	1.319
		Phonological-Low	.275	.448	1.000	-.931	1.481
	Phonological-Low	Executive-High	-.348	.469	1.000	-1.612	.916
		Executive-Low	-.124	.455	1.000	-1.351	1.103
		Phonological-High	-.275	.448	1.000	-1.481	.931
Productive Vocab Posttest	Executive-High	Executive-Low	1.630	1.296	1.000	-5.122	1.862
		Phonological-High	.227	1.274	1.000	-3.206	3.660
		Phonological-Low	-1.435	1.335	1.000	-5.032	2.163
	Executive-Low	Executive-High	-1.630	1.296	1.000	-1.862	5.122
		Phonological-High	-1.857	1.233	.812	-1.465	5.180
		Phonological-Low	.196	1.296	1.000	-3.296	3.688
	Phonological-High	Executive-High	-.227	1.274	1.000	-3.660	3.206
		Executive-Low	1.857	1.233	.812	-5.180	1.465
		Phonological-Low	1.661	1.274	1.000	-5.094	1.771
	Phonological-Low	Executive-High	-1.435	1.335	1.000	-2.163	5.032
		Executive-Low	-.196	1.296	1.000	-3.688	3.296
		Phonological-High	-1.661	1.274	1.000	-1.771	5.094

Dependent Variable	(I) WM	(J) WM	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a]	
						Lower Bound	Upper Bound
Productive Vocab Delayed Post-test	Executive-High	Executive-Low	-1.844	1.215	.794	-5.119	1.430
		Phonological-High	.081	1.195	1.000	-3.300	3.138
		Phonological-Low	1.522	1.252	1.000	-4.895	1.851
	Executive-Low	Executive-High	-1.844	1.215	.794	-1.430	5.119
		Phonological-High	-1.764	1.156	.783	-1.352	4.879
		Phonological-Low	.323	1.215	1.000	-2.952	3.597
	Phonological-High	Executive-High	-.081	1.195	1.000	-3.138	3.300
		Executive-Low	1.764	1.156	.783	-4.879	1.352
		Phonological-Low	1.441	1.195	1.000	-4.660	1.778
	Phonological-Low	Executive-High	-1.522	1.252	1.000	-1.851	4.895
		Executive-Low	-.323	1.215	1.000	-3.597	2.952
		Phonological-High	-1.441	1.195	1.000	-1.778	4.660

According to Table 6, the difference between all aspects of WM and receptive and productive lexical gain was insignificant at baseline ($p > 0.05$). However, on the post-test of vocabulary knowledge, executive-high WM learners outperformed executive-low, phonological-high, and phonological-low WM learners. Similarly, on the post-test of vocabulary knowledge, executive-high WM learners outperformed executive-low, phonological-high, and phonological-low WM learners. This pattern persisted in the results of the delayed post-test. Notwithstanding, attention needs to be paid that in none of the cases the difference between different aspects of WM was significant as $p > 0.05$.

The third research question was concerned with understanding how language proficiency affects receptive and productive lexical gain. To measure this, we needed to run a statistical test of significance. According to Pallant (2020), when an independent variable's effect on more than one dependent variable is investigated, the conduction of MANOVA is warranted. However, before doing this, we needed to ensure the normality assumption. Thus, a K-S test was run whose results substantiated the assumption of normality ($p > 0.05$).

Table 7.
Descriptive Statistics

	Proficiency	Mean	Std. Deviation	N
Receptive Vocab Pre-test	Beginner	3.700	1.454	20
	Lower-intermediate	3.954	1.554	44
	Intermediate	3.722	1.596	36
	Total	3.820	1.539	100
Receptive Vocab Post-test	Beginner	6.200	2.041	20
	Lower-intermediate	6.916	4.249	44
	Intermediate	10.113	5.261	36
	Total	8.180	4.632	100
Receptive Vocab Delayed Post-test	Beginner	5.750	1.943	20
	Lower-intermediate	6.638	3.961	44
	Intermediate	9.727	4.876	36
	Total	7.820	4.351	100
Productive Vocab Pre-test	Beginner	3.300	1.490	20
	Lower-intermediate	3.522	1.649	44
	Intermediate	3.444	1.557	36
	Total	3.450	1.572	100
Productive Vocab Posttest	Beginner	5.500	1.572	20
	Lower-intermediate	6.722	4.368	44
	Intermediate	9.181	5.196	36
	Total	7.560	4.537	100
Productive Vocab Delayed Post-test	Beginner	5.200	1.472	20
	Lower-intermediate	6.305	4.051	44
	Intermediate	8.840	4.856	36
	Total	7.200	4.264	100

According to Table 7, learners with different proficiency levels performed similarly on both pre-tests. However, on the immediate receptive post-test, intermediate learners outperformed other learners (Intermediate learners' mean = 10.113, lower-intermediate learners' mean = 6.916, beginner learners' mean = 6.200). Similarly, on the immediate productive post-test, intermediate learners outperformed other learners (Intermediate learners' mean = 9.181, lower-intermediate learners' mean = 6.722, beginner learners' mean = 5.500). Likewise, on the delayed receptive post-test, the intermediate learners' mean was higher (Intermediate learners' mean = 9.727, lower-intermediate learners' mean = 6.638, beginner learners' mean = 5.750). On the delayed productive post-test, the intermediate learners' mean was higher (Intermediate learners' mean = 8.840, lower-intermediate learners' mean = 6.305, beginner learners' mean = 5.200). However, the fault with descriptive statistics is that it does not show whether the difference between groups is significant statistically. To this end, pairwise comparisons need to be conducted.

Table 8.
Pairwise Comparisons of language proficiency effect on receptive and productive lexical gain

Dependent Variable	(I) Proficiency	(J) Proficiency	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
						Lower Bound	Upper Bound
Receptive Vocab Pre-test	Beginner	Intermediate	-.255	.418	1.000	-1.274	.764
		Lower-intermediate	-.022	.433	1.000	-1.076	1.032
	Intermediate	Beginner	.255	.418	1.000	-.764	1.274
		Lower-intermediate	.232	.349	1.000	-.617	1.081
	Lower-intermediate	Beginner	.022	.433	1.000	-1.032	1.076
		Intermediate	-.232	.349	1.000	-1.081	.617
Receptive Vocab Post-test	Beginner	Intermediate	-3.914*	1.170	.003	-6.763	-1.064
		Lower-intermediate	-.717	1.209	1.000	-3.663	2.230
	Intermediate	Beginner	3.914*	1.170	.003	1.064	6.763
		Lower-intermediate	3.197*	.975	.004	.823	5.571
	Lower-intermediate	Beginner	.717	1.209	1.000	-2.230	3.663
		Intermediate	-3.197*	.975	.004	-5.571	-.823
Receptive Vocab Delayed Post-test	Beginner	Intermediate	-3.977*	1.088	.001	-6.628	-1.327
		Lower-intermediate	-.889	1.125	1.000	-3.630	1.852
	Intermediate	Beginner	3.977*	1.088	.001	1.327	6.628
		Lower-intermediate	3.088*	.907	.003	.880	5.297
	Lower-intermediate	Beginner	.889	1.125	1.000	-1.852	3.630
		Intermediate	-3.088*	.907	.003	-5.297	-.880
Productive Vocab Pre-test	Beginner	Intermediate	-.223	.428	1.000	-1.265	.819
		Lower-intermediate	-.144	.442	1.000	-1.222	.933
	Intermediate	Beginner	.223	.428	1.000	-.819	1.265
		Lower-intermediate	.078	.356	1.000	-.790	.947
	Lower-intermediate	Beginner	.144	.442	1.000	-.933	1.222
		Intermediate	-.078	.356	1.000	-.947	.790
Productive Vocab Posttest	Beginner	Intermediate	-3.682*	1.166	.006	-6.522	-.842
		Lower-intermediate	-1.222	1.206	.940	-4.159	1.715
	Intermediate	Beginner	3.682*	1.166	.006	.842	6.522
		Lower-intermediate	2.460*	.972	.039	.093	4.826
	Lower-intermediate	Beginner	1.222	1.206	.940	-1.715	4.159
		Intermediate	-2.460*	.972	.039	-4.826	-.093
Productive Vocab Delayed Post-test	Beginner	Intermediate	-3.641*	1.086	.003	-6.286	-.995
		Lower-intermediate	-1.106	1.123	.982	-3.841	1.630
	Intermediate	Beginner	3.641*	1.086	.003	.995	6.286

Dependent Variable	(I) Proficiency	(J) Proficiency	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
						Lower Bound	Upper Bound
		Lower-intermediate	2.535*	.905	.018	.331	4.740
	Lower-intermediate	Beginner	1.106	1.123	.982	-1.630	3.841
		Intermediate	-2.535*	.905	.018	-4.740	-.331

According to the table, on pre-tests, the difference between learners with different proficiency levels was not significant ($p > 0.05$). However, the difference was significant on receptive post-test, with intermediate learners outperforming beginner and lower-intermediate learners ($p < 0.05$). However, the difference between lower-intermediate and beginner learners was insignificant ($p < 0.05$). On the post-test of productive knowledge, the difference between intermediate learners and beginner ones was significant ($p < 0.05$), with intermediate learners outperforming lower-intermediate ones. However, the difference between lower-intermediate learners and beginner participants was not significant ($p > 0.05$). The difference was significant on the delayed post-test of receptive knowledge, with intermediate learners outperforming beginner and lower-intermediate learners ($p < 0.05$). However, the difference between lower-intermediate learners with beginner learners was not significant. On the delayed post-test of productive knowledge, the difference between intermediate learners and beginner ones was significant ($p < 0.05$), with intermediate learners outperforming lower-intermediate ones. However, the difference between lower-intermediate learners and beginner participants was not significant ($p > 0.05$).

The results of pairwise comparisons presented above indicated that at baseline, learners' scores were low, but learners with higher WM and language proficiency outperformed themselves on the immediate and delayed post-tests ($p < 0.05$). For a full analysis, please see above.

Discussion

It was found that participants who were engaged in tasks involving blank-fill and production tasks exhibited significantly greater gains in both receptive and productive vocabulary knowledge. Moreover, the interaction of WMC and language proficiency played a crucial role in shaping these outcomes. The results could be attributed to the cognitive mechanisms at play during vocabulary learning. Blank-fill and production tasks create a deeper level of engagement, fostering better retention and recall of lexical items. The higher involvement load required by such tasks may enhance learners' attention, motivation, and generation processes (Teng & Zhang, 2021), facilitating the transfer of vocabulary knowledge from short-term to long-term memory (Teng, 2022). Additionally, participants with high WMC demonstrated improved vocabulary acquisition, potentially due to their ability to process and organize linguistic information more efficiently (Baddeley, 2003; 2017).

The novelty of our research lies in its comprehensive approach, considering not only the word-focused task effects on vocabulary learning but also incorporating the influences of WMC and language proficiency. By exploring these multifaceted factors simultaneously, we provide a more nuanced understanding of how vocabulary acquisition occurs. This study extends beyond the traditional focus on the ILH, shedding light on the interplay between cognitive abilities and instructional techniques. As such, our research offers a valuable contribution to the field of language learning, guiding educators and researchers toward more effective pedagogical strategies tailored to diverse learners' characteristics. Additionally, we delved into the impact of two distinct aspects of WM. Our findings revealed that executive WM emerged as a more influential factor in shaping vocabulary learning outcomes. Learners with stronger executive WMC displayed significantly better results. However, while phonological short-term WM did exhibit an effect on vocabulary learning, the difference it made was not statistically significant.

Our findings support the assertions of Laufer and Hulstijn (2001) regarding the effectiveness of vocabulary tasks that engage learners actively. The results indicate that participants exposed to blank-fill and production tasks demonstrated greater gains in vocabulary knowledge compared to those engaged solely in reading. This underscores the significance of interactive and contextualized vocabulary instruction, a concept central to the ILH. In these tasks (blank-fill and especially production tasks), learners not only encounter words but also actively interact with them in meaningful contexts, promoting deeper understanding and retention. Thus, need and search are important factors of ILH in learning new vocabularies.

Our results are in line with Namaziandost et al.'s (2020) examination of high- and low-involvement load tasks. While they found that high involvement load positively contributed to vocabulary development, our study further supports this by demonstrating that word-focused tasks, particularly those involving blank-fill and production tasks, substantially positively impact both receptive and productive lexical gain.

In contrast to Jafari Gohar et al. (2018), our study delved into the nuanced interaction between WMC, language proficiency, and the ILH in the context of vocabulary acquisition. In contrast, their research focused on the ILH and technique feature analysis's predictive power. Our study explored how individual differences in cognitive capacities and language competence influence vocabulary learning outcomes. Moreover, our investigation extends beyond a single framework to encompass diverse experimental conditions. This comprehensive approach reveals that the efficacy of the ILH varies across conditions and underscores the importance of considering task variations and individual characteristics when optimizing foreign language vocabulary instruction.

In contrast to the study by Teng and Zhang (2021), our research offers a distinct perspective on the ILH by examining its implications within the context of WMC and language proficiency. While they focused on metacognitive aspects and task-based vocabulary learning, our study investigated the multifaceted interplay of ILH, WM, and language competence in lexical gain. They concentrated on different work conditions, revealing varying degrees of success in vocabulary learning outcomes. In contrast, our study encompasses a broader range of task conditions, shedding light on the nuanced relationship between ILH and lexical gain under diverse conditions. Furthermore, we delved into the roles of WM and language proficiency, enhancing our understanding of the intricate mechanisms underlying foreign language vocabulary instruction.

The study conducted by Ansarian and Kazemipour Khabbazi (2021) offers insight that challenges the strict adherence to the ILH. While our study finds support for the effectiveness of ILH on vocabulary acquisition, we also introduce the complexity of WMC and language proficiency, indicating that these factors interact with involvement load to shape vocabulary learning outcomes.

Our study differs from Teng's (2022) research in several key aspects. While Teng focused on investigating how word-focused conditions influenced lexical gain, our study extends beyond this by examining different variables' effects on foreign vocabulary learning. Teng's work primarily delved into the impact of different word-focused exercises, whereas, our research encompasses a broader spectrum, considering three distinct task conditions and how different aspects of WM affect lexical gain. Additionally, we investigated the effects of distinct aspects of WM, while Teng's study lacked this.

Regarding the role of WMC in lexical gain, our findings align closely with Baddeley's (2017) insights. We observed a consistent trend where participants with higher WMC consistently outstripped their lower WMC peers on both immediate and delayed post-tests. This pattern strongly supports the argument that cognitive resources, encapsulated within the domain of WM, hold a pivotal role in the retention and recall of vocabulary (Baddeley, 2017). Additionally, this finding is in contrast with Crossley and Kim (2019), and Kormos and Trebits (2011), who failed to find a link between WMC and language learning.

Regarding the potential of different aspects of WM in receptive and productive lexical gain, our investigation attempted to provide a more nuanced understanding of phonological short-term and executive WM's impact on lexical gain, which, according to Teng (2022), was unexplored. The results demonstrated that both executive and phonological WM exerted influences on lexical gain outcomes. Notably, learners with higher executive WM exhibited slightly better performance compared to those with high phonological WM, although this different did not reach statistical significance. This nuanced exploration highlights the multifaceted nature of WM's contribution to lexical gain and suggests that both executive and phonological components play roles in shaping the complexities of lexical retention and recall in language learning.

Regarding the potential of language proficiency in lexical gain, in line with previous research by Zareva et al. (2005), Nizonkiza (2011), and Tilfarlioglu and Bozgeyik (2012), our study affirmed the substantial impact of language proficiency on receptive and productive lexical gain. Intermediate-level learners consistently outperformed beginners and lower-intermediate peers, raising questions about the mechanisms behind this phenomenon.

Intermediate-level learners benefit from a stronger foundational understanding of grammar and contextual language usage. Their exposure to the language is typically more extensive, resulting in increased vocabulary encounters. This, coupled with their enhanced confidence, contributes to their superior lexical gain. Recognizing these advantages reinforces the importance of tailored instruction to proficiency levels, enhancing vocabulary learning outcomes and overall language proficiency.

Conclusion

In conclusion, we examined the impact of word-focused tasks, WMC, and language proficiency on receptive and productive lexical gain. The findings reveal several key conclusions:

1. **Word-focused tasks enhance vocabulary learning:** Participants engaged in tasks involving need and search demonstrated significantly greater gains in both receptive and productive lexical gain. These tasks create a deeper engagement level, fostering better retention and recall of lexical items, and are thus recommended for vocabulary instruction.
2. **Working memory matters:** WMC plays a crucial role in shaping learning outcomes. Participants with higher WMC consistently outperformed their peers with lower WMC on both immediate and delayed post-tests. This highlights the importance of learners' cognitive resources, particularly their executive WM, in vocabulary retention and recall.
3. **Higher language proficiency results in more lexical gain:** Intermediate-level learners exhibited more robust receptive and productive vocabulary knowledge compared to beginners and lower-intermediate ones. This underscores the importance of tailoring vocabulary instruction to learners' language proficiency levels.
4. **Nuanced Interplay of Factors:** The study's novelty lies in its comprehensive approach, considering the effects of different word-focused tasks, WMC, and language proficiency in one shot. It extends beyond the traditional focus on the ILH, showing that the efficacy of the ILH varies across conditions and is influenced by individual differences in cognitive capacities and language competence.

In summary, this research contributes to understanding foreign lexical gain by highlighting the interplay of various factors and offering practical guidance for educators and policymakers in enhancing language instruction and materials development. Further research in this domain can continue to refine our understanding of these intricate processes.

The findings have significant implications for language teaching and learning. Our research underscores the effectiveness of need and search in enhancing both receptive and productive lexical gain. These tasks promote deeper engagement and foster better retention and recall of lexical items, aligning with the cognitive mechanisms involved in vocabulary learning. Moreover, our study highlights the importance of considering individual differences in language instruction. Learners with higher WMC demonstrated superior lexical gain, emphasizing the role of cognitive resources in the process. Additionally, language proficiency levels influence vocabulary outcomes, advocating for tailored instruction based on learners' language abilities.

The implications for teachers are profound. Educators can harness the insights from this research to craft vocabulary instruction that maximizes student learning. By incorporating interactive tasks with higher involvement loads, such as blank-fill and production tasks, teachers can enhance their students' vocabulary development. Moreover, recognizing the role of language proficiency in lexical gain underscores the importance of adapting instruction to individual student levels. Teachers should tailor their teaching materials and methods to match the proficiency level of their learners, promoting more effective and efficient lexical gain. In a nutshell, this study empowers teachers with evidence-based strategies to facilitate lexical gain and optimize language learning in their classrooms.

Furthermore, our study provides nuanced and actionable insights that can significantly impact language education policies, syllabus design, and materials development. For policymakers, our findings underscore the importance of incorporating a variety of word-focused tasks into language learning programs. Specifically, we recommend the integration of blank-fill and, especially, production tasks, as revealed by our research, to yield the greatest gains in lexical gain. Policymakers should advocate for curricular flexibility to accommodate these task variations.

Syllabus designers, building on our results, should consider tailoring language instruction to learners' cognitive abilities. Recognizing that WMC plays a pivotal role in lexical gain, syllabi can be designed to include tasks that enhance learners' WMC. This approach can significantly contribute to vocabulary development, as our study demonstrated that participants with higher WMC consistently outstripped their peers on both immediate and delayed post-tests.

Materials developers can utilize our research to create more effective and engaging resources that align with cognitive processes involved in lexical gain. Materials should not only emphasize need and search but also provide ample opportunities for learners to train their WM through targeted exercises. By integrating these findings into material development, we can enhance lexical gain.

In short, our study offers detailed and practical recommendations for teachers, policymakers, syllabus designers, and materials developers. These recommendations, grounded in our study, advocate for curricular flexibility, tailored instruction, and materials that align with the cognitive mechanisms involved in lexical gain. This comprehensive approach holds the potential to significantly improve language education outcomes.

However, this study is not without limitations. Our research concentrated on a specific group of participants and employed controlled tasks within a controlled environment, potentially limiting the generalizability of the results to broader language learning contexts. Furthermore, while we explored various aspects of WM, further investigation into specific components like the phonological loop and central executive could offer deeper insights. Further studies might consider diverse learner populations and more ecologically valid settings to ascertain the wider applicability of our findings. Additionally, examining the transfer of vocabulary knowledge across different language skills and exploring the long-term effects of the observed cognitive processes could provide a more comprehensive understanding of effective language instruction.

Acknowledgments
We would like to thank the editorial team of TESL Quarterly for granting us the opportunity to submit and publish the current synthesis. We would also like to express our appreciation to the anonymous reviewers for their careful, detailed reading of our manuscript and their many insightful comments and suggestions.

Declaration of conflicting interests
The authors declare no potential conflicts of interest concerning the research, authorship, and/or publication of this article.

Funding
The authors received no financial support for this article's research, authorship, and/or publication.

References

Ansarin, A. A., & Kazemipour Khabbazi, S. (2021). Task-induced involvement load and working memory: Effects on active and passive vocabulary knowledge of EFL learners in a multimedia learning environment. *Eurasian Journal of Applied Linguistics, 7*(1), 277-302. https://doi.org/10.32601/ejal.911288
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. (2017). Modularity, working memory and language acquisition. *Second Language Research, 33*, 299–311. https://doi.org/10.1177/0267658317709852
Crossley, S. A., & Kim, Y. (2019). Text integration and speaking proficiency: Linguistic, individual differences and strategy use considerations. *Language Assessment Quarterly, 16*(2), 217-235. https://doi.org/10.1080/15434303.2019.1628239
Ellis, R. (2015). *Understanding Second Language Acquisition* (2nd ed.). Oxford University Press.
Fazilatfar, A. M., Ghorbani, S., & Samavarchi, L. (2011). The effect of standard and reversed subtitling versus no subtitling mode on L2 vocabulary learning. *Teaching English as a*

Second Language Quarterly (Formerly Journal of Teaching Language Skills), 30(1), 43-64. https://doi.org/10.22099/jtls.2012.393

Harmer, J. (2001). *The practice of English language teaching*. Pearson.

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

Jafari Gohar, M., Rahmanian, M., & Soleimani, H. (2018). Technique feature analysis or involvement load hypothesis: Estimating their predictive power in vocabulary learning. *Journal of psycholinguistic research, 47*(4), 859-869. https://doi.org/10.1007/s10936-018-9568-5

Hayati, M., & Shahriari, A. (2012). The impact of L1 equivalents versus context on vocabulary recall of pre-university EFL students. *Teaching English as a Second Language Quarterly (Formerly Journal of Teaching Language Skills), 29*(3), 27-51. https://doi.org/10.22099/jtls.2012.414

Kaivanpanah, S., Alavi, S. M., & Ravandpour, A. (2020). The effect of input-based and output-based tasks with different and identical involvement loads on Iranian EFL learners' incidental vocabulary learning. *Cogent Psychology, 7*(1), 1731223. http://doi.org/10.1080/23311908.2020.1731223

Kivrak, C., & Gökmen, D. U. (2019). The effects of task-induced involvement load and input modality on incidental vocabulary learning. *Journal of Language and Linguistic Studies, 15*(4), 1355-1375. https://doi.org/10.17263/jlls.668478

Kormos, J., & Trebits, A. (2011). Working memory capacity and narrative task performance. In
P. Robinson (Ed.), *Second language task complexity: Researching the cognition hypothesis of language learning and performance* (pp. 267–289). John Benjamins.

Laufer, B. (2003). Vocabulary acquisition in a second language: Do learners really acquire most vocabulary by reading? Some empirical evidence. *Canadian Modern Language Review, 59*(4), 567-587. http://dx.doi.org/10.3138/cmlr.59.4.567

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

Miyake, A., & Friedman, N. P. (1998). Individual differences in second language proficiency:
Working memory as language aptitude. In A. F. Healy (Ed.), *Foreign language learning* (pp. 339–364). Lawrence Erlbaum Associates.

Namaziandost, E., Hosseini, E., & Utomo, D. W. (2020). A comparative effect of high involvement load versus lack of involvement load on vocabulary learning among Iranian sophomore EFL learners. *Cogent Arts & Humanities, 7*(1), 1715525. https://doi.org/10.1080/23311983.2020.1715525

Nassaji, H., & Hu, H. C. M. (2012). The relationship between task-induced involvement load and learning new words from context. *International Review of Applied Linguistics in Language Teaching, 50*(1), 69-86. http://doi.org/10.1515/iral-2012-0003

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge University Press.

Nation, I. S. P., & Webb, S. A. (2011). *Researching and analyzing vocabulary*. Cengage Learning.

Nizonkiza, D. (2011). The relationship between lexical competence, collocational competence, and second language proficiency. *English Text Construction*, *4*(1), 113-145. https://doi.org/10.1075/etc.4.1.06niz

Pallant, J. (2020). *SPSS survival manual: A step-by-step guide to data analysis using IBM SPSS*. McGraw-Hill Education.

Paribakht, T., & Wesche, M. (1996). Vocabulary enhancement activities and reading for meaning in second language vocabulary acquisition. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition: A rationale for pedagogy* (pp.174-200). Cambridge University Press. https://doi.org/10.1017/CBO9781139524643.013

Schmitt, N. (2000). *Vocabulary in Language Teaching*. Cambridge University Press.

Shahnazari, M. (2013). The development of a Persian reading span test for the measure of L1 Persian EFL learners' working memory capacity. *Applied Research on English Language*, *2*(2), 107-116. https://doi.org/10.22108/are.2013.15473

Teng, M. F. (2022). The Roles of Second-Language Proficiency Level and Working Memory on Vocabulary Learning from Word-Focused Exercises. *RELC Journal*. https://doi.org/10.1177/00336882221102228

Teng, M. F., & Zhang, D. (2021). Task-induced involvement load, vocabulary learning in a foreign language, and their association with metacognition. *Language Teaching Research.* https://doi.org/10.1177/13621688211008798

Thornbury, S. (2002). *How to teach vocabulary*. Pearson Education Limited.

Tilfarlioglu, F. Y., & Bozgeyik, Y. (2012). The relationship between vocabulary learning strategies and vocabulary proficiency of English language learners. *International Journal of Applied Linguistics and English Literature*, *1*(2), 91-100. http://doi.org/10.7575/ijalel.v.1n.2p.91

Wen, Z. E. (2016). *Working memory and second language learning: Toward an integrated approach*. Multilingual Matters.

Wilkins, D. A. (1972). *Linguistics in language teaching*. Edward Arnold.

Yaqubi, B., Rayati, R. A., & Allemzade Gorgi, N. (2012). The involvement load hypothesis and vocabulary learning: The effect of task types and involvement index on L2 vocabulary acquisition. *Teaching English as a Second Language Quarterly (Formerly Journal of Teaching Language Skills)*, *29*(1), 145-163. https://doi.org/10.22099/jtls.2012.404

Zareva, A., Schwanenflugel, P., & Nikolova, Y. (2005). Relationship between lexical competence and language proficiency: Variable sensitivity. *Studies in Second Language Acquisition*, *27*(4), 567-595. https://doi.org/10.1017/S0272263105050254