E-ISSN: 2981-1686

Journal of Studies in Language Learning and Teaching

www.jsllt.yazd.ac.ir

JSLLT, Vol. 3, No. 2, 37–56.

Paper Type: Original Article

Task Modality as a Moderator of Involvement Load in L2 Vocabulary Acquisition and Retention: Evidence from Iranian EFL Learners

Mahdieh Sattar[1,*], Mohammad Golshan[2], Ali Akbar Moeen[3]

[1] Department of English Language, May.C., Islamic Azad University, Maybod, Iran;
[2] Department of English Language, May.C., Islamic Azad University, Maybod, Iran;
[3] Department of English Language, May.C., Islamic Azad University, Maybod, Iran.

Received: 22 June, 2026 Revised: 12 September, 2026 Accepted: 17 September, 2026

Abstract

Although the Involvement Load Hypothesis (ILH) is an influential framework in L2 vocabulary research, relatively few studies have examined the isolated contribution of the Need component while holding other involvement factors constant. Addressing this gap, the present study investigated whether task modality (oral vs. written) moderates the effectiveness of task-induced involvement when operationalized solely through two levels of Need (strong vs. moderate), with Search and Evaluation controlled across conditions. To this end, sixty-four intermediate Iranian EFL learners from four intact classes were assigned at the class level to four experimental conditions (n = 16 per group, age range: 14-16, *M*: 15): composition writing (strong Need), oral storytelling (strong Need), written judgment and rewording (moderate Need), and oral summarization (moderate Need). Receptive vocabulary knowledge was assessed through immediate and delayed posttests administered three weeks apart. A mixed-design ANOVA revealed significant main effects for time, task modality, and Need. Although the interaction between Need and modality was not significant, there emerged a significant three-way interaction among time, Need, and modality, accounting for 21.1% of the variance in retention scores (ηp^2 = .211). Composition writing produced the highest levels of both immediate acquisition and delayed retention. These findings suggest that the effectiveness of task-induced involvement depends not only on the level of Need but also on the modality through which it is enacted, highlighting the pedagogical value of high-Need written tasks for durable vocabulary retention in EFL contexts.

Keywords: Involvement load hypothesis, task modality, incidental vocabulary acquisition, vocabulary retention, EFL context.

I | INTRODUCTION

Vocabulary skills are a key part of communication skills that forms the basis for the learning of a second language (L2). A strong lexicon provides students with the language tools they need to express complex ideas effectively and helps them decode complex verbal and written inputs. According to Nation (2001) and Schmitt (2008), lexical competence is in fact widely acknowledged as the cornerstone for progressing in all primary language modalities, including speaking, listening, reading, and writing. Researchers have continuously looked for pedagogical settings and task designs that maximize the successful acquisition and long-term retention of new lexical items in L2 and foreign language settings, given its crucial role in language learning (Nation, 2013).

Licensee **Journal of Studies in Language Learning and Teaching.** This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-NC) license.

Corresponding Author: Mohammad.Golshan@iau.ac.ir

Cite this article: Sattar, M., Golshan, M., Moeen, A.A., (2026), Task Modality as a Moderator of Involvement Load in L2 Vocabulary Acquisition and Retention: Evidence from Iranian EFL Learners, *Journal of Studies in Language Learning and Teaching*, *3*(2), 37-56. doi: 10.22034/JSLLT.2026.24803.1137.

In the pursuit of maximizing instructional efficiency, task-based instruction has become a major focus in L2 research. This interest stems from the premise that tasks can be designed to create specific cognitive conditions that necessitate deeper mental processing of target words (Ellis et al., 2019). The most influential theoretical framework in this domain is the Involvement Load Hypothesis (ILH), proposed by Laufer and Hulstijn (2001). Grounded in Craik and Lockhart's (1972) Levels of Processing (LoP) framework, the ILH posits that the effectiveness of a task in promoting vocabulary learning is determined by the "involvement load" that it imposes on the learner. This involvement is defined as a cognitive-motivational construct consisting of three components: Need (the motivational dimension), Search, and Evaluation (the cognitive dimensions). The core tenet of the hypothesis is that tasks with a higher involvement load, namely those requiring more intense cognitive engagement, yield superior outcomes in both initial acquisition and subsequent long-term retention (Laufer & Hulstijn, 2001; Yanagisawa & Webb, 2021).

Over the past two decades, an extensive body of empirical research has validated the core tenets of the ILH. Studies have consistently demonstrated that tasks with higher involvement loads such as sentence writing or composition yield greater vocabulary gains than lower-load tasks like reading with marginal glosses or gap-fill exercises (Keating, 2008; Kim, 2008; Webb, 2005). However, a critical synthesis of this literature reveals a distinct methodological pattern: these investigations have operationalized involvement load almost exclusively through literacy-based, written task formats. While this robust body of work has provided invaluable insights into how cognitive effort facilitates lexical retention, it leaves a notable gap regarding how these cognitive mechanisms operate across different modalities of language production. Consequently, it remains unclear whether the predictive power of the ILH remains constant when tasks shift from written to oral mediums.

Importantly, within contemporary Second Language Acquisition (SLA) research, task modality is no longer viewed as a simple binary choice between speaking and writing. Rather, modality is better understood as a continuum of processing conditions that impose varying degrees of communicative pressure, planning time, and working memory demands (Ellis, 2005; Gilabert et al., 2009). Theoretical perspectives on L2 production suggest that oral and written modalities recruit distinct cognitive resources along this continuum. Writing typically resides on the end of the spectrum that allows for reflection, revision, and a slower pace of lexical retrieval. This decontextualized and deliberate nature of writing may provide learners with the necessary additional processing time to focus on form-meaning mappings, thereby potentially enhancing the effects of a high involvement load.

Conversely, spoken production leans toward the opposite end of the continuum, characterized by real-time processing, fleeting input, and higher communicative pressure. The impact of these oral processing conditions on vocabulary acquisition presents competing hypotheses. In one respect, according to Skehan's (1998) Limited Attentional Capacity Model, learners have finite cognitive resources; in the heat of oral interaction, the heavy demands of maintaining fluency and managing real-time message delivery may deplete the attentional resources available for deep lexical processing, potentially diminishing the benefits of a high involvement load. On the other hand, competing perspectives, such as the output hypothesis (Swain, 2005) and the practice paradigms of the practice of the language, suggest that the oral method may be uniquely suited to facilitate vocabulary acquisition. The dynamic nature of speech requires rapid lexical retrieval, active articulation and often negotiation of meaning, which may produce powerful phonetic and motor memory imprints. In this sense, the immediate communicative pressure of oral tasks may actually make students more aware of language deficits and reinforce form-meaning mapping in ways that are not possible with written tasks.

Investigating the interplay of task modality and involvement load is essential for several reasons. First, it addresses the "modality-blind" nature of the original ILH, potentially refining the hypothesis by identifying the boundary conditions under which it operates most effectively. Second, it offers much-needed empirical evidence for EFL practitioners who must decide how to balance oral and written tasks to maximize vocabulary gains. By examining whether the benefits of involvement load are contingent upon the processing conditions imposed by the mode of production, this study seeks to provide a more nuanced account of how task design interacts with the cognitive architecture of the L2 learner.

The primary objective of the present study is to examine whether task modality (oral vs. written) moderates L2 vocabulary learning under varying levels of task-induced Need. Following the

methodology suggested by Laufer and Hulstijn (2001), the study compares tasks involving strong versus moderate Need while maintaining the other involvement components (Search and Evaluation) at a constant level. This design allows for a precise isolation of how the "Need" component interacts with output mode to shape immediate vocabulary gains and long-term retention among Iranian EFL learners.

II. LITERATURE REVIEW

2.1. Vocabulary Learning in Task-Based Language Learning Contexts

Tasks-based language learning is a widely used method which is primarily focused on meaningful and goal-oriented language use, but it is also used for non-linguistic purposes (Ellis, 2003; Nunan, 2004). From a different point of view, Long (2015) and Ellis (2018) argue that the tasks in this context are a deliberate activity, involving students in understanding through interpretation, processing or generation of language, and placing the main idea above strict grammatical forms (Ellis, 2003; Nunan, 2004).Recent advances in TBLT highlight that specific elements of the design of the task, such as cognitive load, conditions and complexity, determine how learners interact with the input and output of the task, which in turn influences their language development (Ellis et al., 2019). These shifts provide a solid basis for exploring how different task settings could differentiate the acquisition of vocabulary L2.

Vocabulary is a key component of L2 competence and is the basis for the use of text, ideas and interaction (Nation, 2001; Schmitt, 2010, 2019). From a pedagogical point of view, task-based contexts can support the growth of vocabulary by facilitating the integration of different types of lexical exposure, both intentional and incidental, during purposeful communication (Webb et al., 2020; Webb & Nation, 2017). Of course, tasks that require expanding the text to interpret, reconstruct, or solve problems about meaning tend to create a high degree of target word-finding, chance to elaborate, and forced output in the hope of accelerating progress. But learners should not believe that more time spent on tasks always means more new vocabulary to learn: vocabulary acquisition should not be seen as a one-way process, where one learns many things, but as a step-by-step process, which can be distinguished between different types of learning (Nation, 2013; Schmitt, 2010). Engagement in this task can result in: recognition (storage of a new word and its root meaning in memory), partial recognition (acquiring a more accurate understanding) deeper recognition (acquiring a knowledge of a word sufficient to use it in speech or writing). Only with these deeper types of processing can there be deep knowledge (receptive and productive) and long-term preservation provided that the tasks are carefully designed to allow for varying levels of lexical involvement (Schmitt, 2019; Webb et al., 2020).

Emphasizing student interaction with new words during tasks aligns well with the Involvement Load Hypothesis (ILH) introduced by Laufer and Hulstijn (2001). This hypothesis establishes a model based on cognition and motivation to forecast how well words are learned depending on what the activity requires. The core premise of the ILH is that the success of an activity designed for word acquisition relies on how much it triggers three specific mental states, including Need, Search, and Evaluation, while the student processes the new words. Following this, various practical studies have shown that activities demanding greater cognitive involvement generally lead to stronger retention of words (e.g., Kim, 2008; Zou, 2017; Yanagisawa & Webb, 2021). This indicates that the way a task is structured heavily influences lexical development. Nevertheless, a significant portion of these studies has primarily contrasted different task categories based on their cognitive demands, largely overlooking how particular execution factors, like whether a task is performed orally or in writing, might alter these involvement factors. Expanding on previous research, this paper explores the acquisition of vocabulary within task-driven environments utilizing the ILH framework, specifically focusing on how oral versus written execution could uniquely activate the processes of Need, Search, and Evaluation in students. Furthermore, while TBLT emphasizes the importance of task features, it often lacks a specific mechanism to predict lexical success. This is where the Involvement Load Hypothesis (ILH) provides a necessary cognitive framework, bridging the gap between general task design and specific lexical processing.

2.2. The Involvement Load Hypothesis

Introduced by Laufer and Hulstijn (2001), the Involvement Load Hypothesis (ILH) stands as a highly prominent cognitive model for anticipating the success of word acquisition within formal L2 education. A fundamental claim of the ILH is that we can gauge how well a task promotes word learning by looking at its capacity to generate three distinct types of engagement, namely Need, Search, and Evaluation, as

students interact with the intended vocabulary. Based on this hypothesis, it is assumed that greater involvement in mental work with the target vocabulary will produce clearer impressions in the memory and, consequently, a longer lasting memory.

The first part of the hypothesis, Need, describes the driving force or pressure in the mind that drives one to say or to recognize a particular word (Laufer and Hulstijn, 2001). The desire to use a specific word may result from a task requiring the production of the word in question externally, or it may simply come from a learner's choice of word to achieve the communicative objective of the interview. Search includes efforts to discover the meaning or spelling of an unknown word, for example. by reference to dictionaries or by asking the native speaker. The final activity is evaluation, a process whereby one compares or manipulates the meaning of the word completion of a task. For example, one of these tasks involves choosing the most appropriate synonym for a given word, or putting the word in context, or estimating the appropriateness of a word for a given sentence. If this theory is correct, then tasks that increase the demand for these three areas of focus will lead learners to acquire more vocabulary over time (Laufer & Hulstijn, 2001). Table 1 provides a detailed description of these components along with examples of their operationalization in task requirements.

To make this hypothesis operational, Laufer and Hulstijn (2001) introduced rules for coding tasks. They assign numbers to each component according to the presence or absence of that component at that level. Generally, the Search is assigned a value of 0 means it is missing, and 1 means it is present. Evaluation and Need reflect different levels of Need and Evaluation, so these categories are divided into three levels 0 for absent Need, 1 for moderate Need (e.g., need to be compared to one item), 2 for a strong Need (internalized need) or an Evaluation (complex integration of contexts). Together, these make up the Involvement Load Index, a quantifiable measure that researchers can use to compare the potential contribution of different tasks to vocabulary learning.

Although the Involvement Load Hypothesis conceptualizes involvement as the combined effect of Need, Search, and Evaluation, empirical studies often manipulate only one component while keeping the others constant in order to isolate its contribution to vocabulary learning. Among these components, Need has been considered particularly influential because it represents the motivational drive that pushes learners to process lexical items during task completion. Tasks that impose a strong Need, such as requiring the obligatory use of target words, are expected to trigger deeper engagement with lexical items than tasks in which the use of such words is only moderately required.

More interesting is the fact that meta-analytic evidence and recent reviews show that the ILH is still a valid theoretical system (Webb et al., 2019; Hulstijn, 2019; Schmitt, 2019). However, various researchers have pointed out that the operationalization of the ILH is still relatively broad, so the first limitation that should be noted is the necessity of making a distinction between the burden of participation and the burden of general cognitive work. The load of involvement is actually about positive, task-related lexical activation (e.g., the German cognitive load) is the cognitive load that the learner needs to process the target words, whereas the general cognitive load is essentially all the main tasks that the working memory has to perform, especially when the task is performed (external load). According to Baralt & Gmez (2017), where the mode of execution (e.g., the procedure for the execution of the sentence) is concerned. In real time, the additional external pressure on the learner's processing capacity may prevent him from performing the cognitively necessary actions of finding and assessing needs, making the high workload counterproductive and hence ineffective.

Second, the operationalization of ILH must explicitly acknowledge individual learner differences and prior knowledge. A task's theoretical involvement index is not absolute; it interacts dynamically with a learner's baseline proficiency. For example, a task designed to elicit 'Search' (e.g., using a dictionary) will not actualize this component if a learner already possesses prior knowledge of the target word. Similarly, working memory capacity and L2 proficiency heavily mediate how learners perceive 'Need' and execute 'Evaluation' (Zou et al., 2021). Therefore, a task with a uniform involvement index may impose vastly different actual involvement loads depending on the learner's profile.

Table 1. Components of the Involvement Load Hypothesis (adopted from Laufer and Hulstijn, 2001)

Component	Description	Example of task requirement
Need	The motivational requirement to use or understand a lexical item which may be externally imposed or internally generated	Learners must use target words to complete a task or convey meaning.
Search	The attempt to find the meaning or form of an unknown word	Looking up a word in a dictionary or asking peers for meaning
Evaluation	Comparing a word with alternatives or deciding whether it appropriately fits a context	Choosing the best word to complete a sentence or integrating a word into a text

These development highlights the need to look at task variables beyond the types of tasks that affect the load and vocabulary acquisition. For example, there is a difference between talking about a task and writing about it. The learner who uses the task orally versus in writing may be influenced differently by the three components of Need, Search and Evaluation. Oral tasks, for example, are often associated with simultaneous information processing and semantic bargaining, which can be interpreted as reinforcing Need. However, written tasks provide the opportunity for a reflection Search and a more detailed analysis of lexical possibilities.

2.3. Empirical Studies on the Involvement Load Hypothesis and Vocabulary Learning

Since the introduction of the Involvement Load Hypothesis (ILH) by Laufer and Hulstijn (2001), a substantial body of empirical research has investigated the relationship between task-induced involvement and vocabulary acquisition. While early reviews often presented a chronological narrative, a more systematic examination of the literature reveals that the predictive power of the ILH is heavily moderated by several key methodological dimensions. Specifically, previous findings can be systematically organized by examining (1) task type, (2) the timing of outcomes (immediate vs. delayed), (3) test type, and (4) learner age and proficiency.

2.3.1. Task Type and Immediate vs. Delayed Outcomes

The most consistent support for the ILH results from studies in which task types were manipulated with a particular emphasis on contrasting receptive tasks (e.g., reading with highlights and production tasks (e.g., writing of sentences and composing). These early papers, by Hulstijn and Laufer (2001), Keating (2008) and Kim (2008), show that production tasks, which include a greater need for Evaluation, involve more complex thinking and processing, using or creating new linguistic elements. Interestingly, while high-involvement load tasks were associated with better and more stable immediate results, the most notable benefit of these types of tasks is revealed in delayed post-test results. In fact, the fact that long-term exposure to these tasks has a positive effect on memory retention is supported by recent published studies, including the detailed meta-analysis of Zou (2017) and Yanagisawa and Webb (2021), which have shown that the generative effort characterized by high-power tasks is effective in protecting against memory decline. In essence, therefore, generative activities where learners must be very hands-on with the new vocabulary (e.g., writing) leave increasingly long-lasting memory traces, which means that the rate of loss of these words after learning is much lower than in cases where the learner's involvement is only passive.

2.3.2. Test Type

The effectiveness of involvement load is also highly sensitive to the *test type* used to measure vocabulary acquisition (e.g., receptive recognition vs. productive recall). Recent meta-analytic and empirical works (Yanagisawa & Webb, 2021) indicate that the benefits of high-load tasks are most pronounced when measured through productive tests (where learners must retrieve and produce the word form or meaning). Studies by Laufer and Rozovski-Roitblat (2011) and Pichette et al. (2012) similarly highlight that tasks involving productive evaluation (e.g., writing-based tasks) align more closely with productive testing formats, whereas tasks with limited lexical decision-making might only show benefits on simpler receptive recognition tests. Therefore, the choice of assessment tool fundamentally shapes whether the benefits of a high involvement index are observable.

2.3.3. Learner Age and Proficiency

Despite the strong empirical support for the ILH, a number of studies have reported mixed results, which are largely attributable to learner-related factors such as age and level of L2 competence. For example,

Martínez-Fernández (2008) and Un-udom (2018) found that tasks with lower levels of involvement sometimes produced comparable or better learning outcomes than those with high levels of involvement. These discrepancies are often explained by the cognitive demands placed on the different skill groups. As highlighted by Hu and Nassaji (2016) and Zou (2017), advanced adult learners have the linguistic resources to cope with highly instrumental tasks (e.g., writing a composition). In contrast, for young learners or students with low proficiency, tasks with an excessive degree of involvement may result in cognitive overload, thereby negating the theoretical benefits of the ILH. In fact, tasks with a moderate degree of involvement may be more facilitating for these learners.

Collectively, this systematic overview suggests that the relationship between the intensity of participation and vocabulary learning is not purely linear, but is moderated by the design of the tasks, the assessment methods and the learner profiles. While the type of task (e.g. reading versus writing) has been extensively studied, another critical variable that can affect the implementation of the components of the engagement is the mode of the task (i.e. whether it is orally or written). To date, only a limited number of studies have directly investigated whether differences in the way students perform tasks affect how they experience and benefit from the inclusion of the components proposed by the ILH, which makes it an important area for further research.

2.4. Task Mode in Vocabulary Learning (Oral vs. Written Tasks)

Within task-based language teaching (TBLT), task mode refers to the modality through which learners perform a task, typically through spoken or written production. Tasks constitute a central mechanism for promoting meaningful language use and encouraging learners to engage in authentic communication (Ellis, 2018; Long, 2015; Skehan, 2018). Because spoken and written tasks involve different cognitive processes and production conditions, they may influence how learners attend to and process lexical items during task performance.

From a cognitive point of view, differences in modalities may affect the way learners allocate attentional resources during performance of a task. Speech production typically involves rapid processing and limited time to reflect, which can limit deliberate Search and Evaluation of the language. Writing, on the other hand, allows students to slow down the process of writing, to revisit the possibilities of language, and to engage in more controlled form-mapping. These modal differences suggest that the effectiveness of task-based engagement may vary according to whether the task is done orally or written.

Further evidence for modality-related differences in lexical processing comes from research examining how production mode influences learners' attention to linguistic form. For example, Adams and Ross-Feldman (2008) found that written task performance provided learners with greater opportunities for noticing and modifying linguistic forms compared to oral interaction. Because writing allows learners to slow down the production process and revisit their output, it often promotes more deliberate attention to vocabulary choice and accuracy. In contrast, spoken interaction tends to prioritize communicative fluency and real-time message delivery, which may reduce the extent to which learners consciously evaluate lexical alternatives. These findings suggest that the modality through which a task is performed can shape the depth of lexical processing, potentially influencing how effectively learners engage with new vocabulary during task completion.

The main difference between these two modes of transport is in their cognitive processing requirements. Speech tasks require students to process and produce the language in real time, often prioritizing fluency and meaning over language accuracy (Ellis, 2018). Due to the pressure of time and the immediacy of interaction, learners tend to concentrate their attention primarily on making sense. By contrast, written assignments typically allow for more planning, monitoring and revision, which allows students to concentrate more on language form and choice of vocabulary (Ellis, 2018; Webb et al., 2020). As a result, writing often encourages more careful verbal and syntactic processing.

Empirical research has provided evidence of these differences. Laufer and Rozovski-Rothblat (2011) reported that written exercises followed by reading tasks led to greater vocabulary increases than verbal activities, suggesting that written work promotes the deeper processing of target words. The TBLT research also shows that written tasks tend to generate more attention to form, while oral tasks often foster interaction and discussion of meaning (Gallardo del Puerto and Martínez Adrián, 2022). Recently, Zhang and Zhang (2022) have found that written tasks provide better vocabulary retention than oral tasks, attributing this advantage to the self-paced nature of writing, which allows students more time to process

information more profoundly and to map out precise form-related meanings. Similarly, Shin (2024) has shown that written modes reduce working memory constraints during the production process, allowing learners to allocate more attentional resources to word-finding, and that reading and writing tasks have a stronger effect on vocabulary acquisition than listening tasks in young learners. But it is important not to idealize writing as being inherently more reflective or more efficient. Depending on the task and the learner's motivation, writing may sometimes result in a shallow or mechanical mapping of form rather than a deep cognitive processing.

Despite these findings, oral and written vocabulary learning are both more effective. First, written vocabulary tasks, by allowing time to plan, offer the possibility of a large-scale lexical search which is not possible in spoken language. Oral vocabulary-learning tasks are, on the other hand, those that stimulate cognition in another and even more important way. As language itself is a means of learning reinforcement, learners are constantly pressed to extract lexemes quickly and immediately from their memory stores when they are asked to speak - this reinforces learning pathways. Speech production is also an important vehicle for the development of communicative language skills; phonological memory is strengthened through pronunciation and interactive learning - a type of learning in which learners identify incompatibilities with the target language due to different mother tongues and resolve them immediately through instant interpretation and feedback. (Ellis, 2018; Webb, 2019).

These mixed findings suggest that the impact of task mode may depend on how tasks direct learners' attention to lexical items and the degree of cognitive engagement they generate. Importantly, the mode through which a task is performed may also interact with the degree of involvement load experienced by learners, as the components of involvement Need, Search, and Evaluation are closely tied to the attentional and cognitive demands that oral and written conditions differentially impose.

Overall, previous research provides substantial—though not entirely consistent—support for the Involvement Load Hypothesis in L2 vocabulary learning. At the same time, task-based studies on modality suggest that oral and written production differ in terms of processing demands, planning opportunities, and attentional allocation. Despite these insights, relatively little research has examined how task modality interacts with specific components of involvement to influence vocabulary learning outcomes. In particular, the extent to which modality may shape the effectiveness of tasks imposing different levels of Need remains largely unexplored. Addressing this gap requires further empirical investigation into whether task modality interacts with varying levels of Need to influence vocabulary learning outcomes among EFL learners.

Building on the Involvement Load Hypothesis and previous research on task modality, the present study aims to examine whether task modality, oral or written, moderates the effects of involvement load, operationalized through different levels of Need, strong and moderate, on second language vocabulary acquisition and retention among Iranian EFL learners. More specifically, the study investigates whether task modality and levels of Need interact to influence learners' immediate vocabulary learning and delayed retention. Thus, the research questions to answer are as follows:

a) Does task modality (oral vs. written) significantly affect the vocabulary acquisition and retention of Iranian EFL learners?

b) Does task modality moderate the effect of task-induced involvement load (operationalized through strong vs. moderate Need) on Iranian EFL learners' vocabulary acquisition and retention?

III. METHODOLOGY

3.1. Participants

The participants of this study consisted of 64 intermediate Iranian EFL learners, comprising 32 males and 32 females. They were selected from four intact classes at a private language institute in Iran. To ensure gender balance across conditions, each experimental group (n = 16) included an equal number of male (n = 8) and female (n = 8) learners. The participants ranged from 14 to 16 years of age ($M = 15.00, SD = 0.72$). All the participants were native speakers of Persian, and their English proficiency level was verified as intermediate based on their scores on the Oxford Placement Test (OPT). Furthermore, to ensure the homogeneity of the groups, a one-way ANOVA was conducted on the OPT scores, which revealed no

significant differences in English proficiency among the four groups at the onset of the study, $F(3,60) = 0.11$, $p = 0.87$. To minimize the potential influence of teacher-related variables, all the four intact classes were taught by the same instructor. This ensured consistency in the delivery of instructions and the overall classroom environment across all the experimental conditions.

3.2. Research Design

Because the study involved pre-existing intact classes, a quasi-experimental design was employed. The study adopted a $2 \times 2 \times 2$mixed factorial design to examine the effects of the task modality and the task-induced Need level on incidental vocabulary learning and retention. The design included two between-subjects factors: Task Modality with two levels, written and spoken, and Need Level with two levels, strong and moderate. Testing Time, with two levels, immediate posttest and delayed posttest, served as the within-subjects factor.

The two levels of task modality and Need were operationalized through four task conditions including a)composition writing, representing the written/strong-Need condition, b) spoken storytelling, representing the spoken/strong-Need condition, c)written judgment and rewording, representing the written/moderate-Need condition, d) and oral summarization, representing the spoken/moderate-Need condition. This design allowed the study to examine not only the main effects of task modality, Need level, and testing time, but also their interaction in shaping vocabulary learning and retention. It is important to note that while the tasks naturally vary in surface genre (e.g., composition/storytelling vs. rewording), this variation was a deliberate methodological choice rather than a confounding variable. To systematically test the Involvement Load Hypothesis (Laufer & Hulstijn, 2001), it was necessary to manipulate the 'Need' component while keeping other components relatively constant across conditions (see Table 2).

As Table 2 illustrates, all tasks required dictionary consultation (Search) and involved contextual integration (Evaluation). However, the tasks deliberately differed in the 'Need' component. The composition and storytelling tasks induced a 'strong Need', because learners had to self-identify lexical gaps and independently select words. Conversely, the rewording and judgment/rewording tasks were selected to operationalize a 'moderate Need', as the target lexical items were preselected and provided by the instructor. Therefore, the distinct task formats served as the essential mechanism to practically implement this specific variation in involvement load. Figure 1 shows the distribution of groups and task conditions.

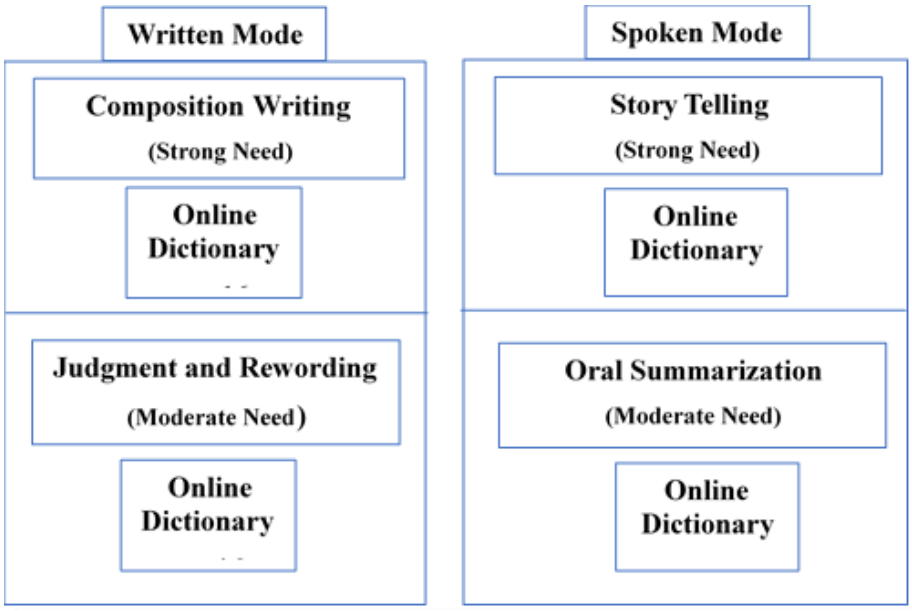


Figure 1. Groups of the study

3.3. Target Words and Instruments

3.3.1. Target Words Selection

To select the target vocabulary, a meticulous screening procedure was conducted to ensure the words represented genuine lexical gaps for the learners. First, based on the gaps explicitly marked as unknown

by participants in the strong-Need conditions, the researchers extracted an initial pool of the 31 most frequently identified words.

According to the ILH (Laufer & Hulstijn, 2001), a strong Need must be internally motivated. Administering a pretest of these 31 words back to the strong-Need groups would have caused pretest sensitization (priming effects), exposing them to the target items beforehand and eliminating the genuine experience of a "lexical gap" during their production tasks. Therefore, a symmetrical pretest across all groups was theoretically impossible.

Instead, to verify that these words were also unfamiliar to the rest of the cohort, the initial 31-word pool was administered as a pretest exclusively to the moderate-Need groups. From this screening test, exactly 20 words that were confirmed to be entirely unknown to the learners in the moderate-Need groups were selected as the final target vocabulary. This asymmetrical, rigorous selection process was a strict theoretical necessity to maintain the construct validity of the strong-Need condition while ensuring target word unfamiliarity across the study (see Figure 2).

3.3.2. Vocabulary Assessments

Vocabulary learning and retention were measured through a pretest, an immediate posttest and a delayed posttest administered three weeks later. Each test was researcher-designed and consisted of 20 multiple-choice items targeting the selected vocabulary introduced in the tasks, with a maximum possible score of 20.

The items were developed to assess the receptive knowledge of the target words in new contexts, with each item presenting four response options (one correct answer and three distractors). To ensure content validity, the items were reviewed by two experienced EFL instructors for alignment with task objectives and appropriate difficulty level. Internal consistency was assessed through test-retest reliability. The Pearson correlation between immediate and delayed posttest scores was $r = 0.91, p < .001$, indicating acceptable stability of the measure over time.The same 20 target words were assessed on both occasions; however, the order of the items and response options was rearranged to reduce potential test familiarity effects. The immediate posttest was used to assess the initial vocabulary gains following the task completion, whereas the delayed posttest was used to measure retention over time. Both tests were scored dichotomously, with one point awarded for each correct answer and zero points for incorrect answers. The testing procedure is illustrated in Figure 2.

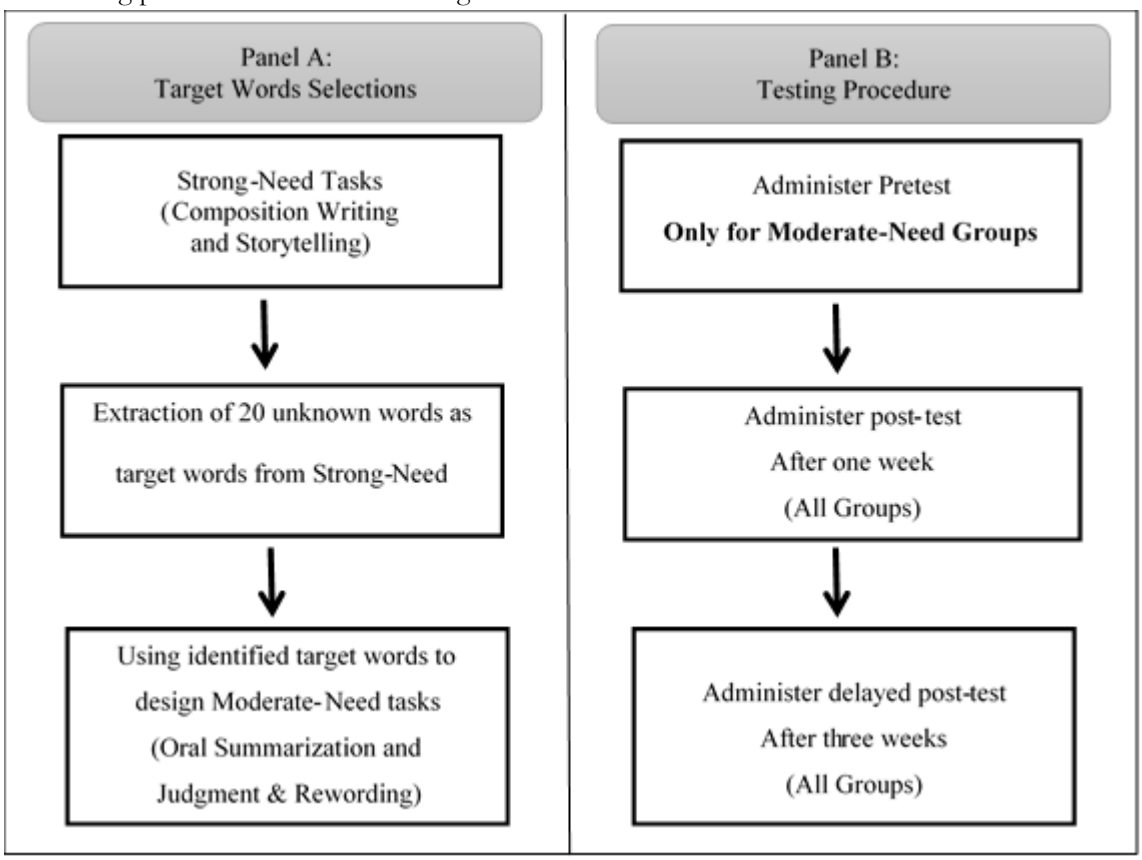


Figure 2. Overview of the Study Design and Experimental Procedure

3.4. Tasks and Operationalization of Need

Four task types were designed based on the Involvement Load Hypothesis (Laufer & Hulstijn, 2001). In this study, the primary manipulated component was Need. Crucially, in line with the original ILH framework, Need was operationalized purely as a motivational dimension (based on the internal versus external source of lexical necessity) rather than a generative cognitive demand. The other two involvement components, Search and Evaluation, were designed to remain as comparable as possible across the task conditions. Table 2 provides a detailed operationalization of how each component was implemented in the task design.

Table 2. Operational coding of task-induced involvement load

Component	ILH definition (Laufer & Hulstijn, 2001)	Strong-Need task (composition & storytelling)	Moderate-Need task (rewording & judgment/rewording)
Need	Internally vs. externally imposed lexical necessity	Learners self-identified lexical gaps during tasks and independently selected words to express intended meanings	Target lexical items were preselected and provided by the instructor.
Search	Effort to find meaning/form	Mandatory bilingual dictionary consultation	Mandatory bilingual dictionary consultation
Evaluation	Contextual comparison and integration	Integration of selected lexical items into original discourse	Contextual evaluation during sentence reformulation

The written composition and spoken storytelling tasks were classified as strong-Need tasks because the necessity for the target words was internally driven; learners had to fill their own lexical gaps in order to express their intended meanings successfully. In these conditions, Need was internally generated through learners' communicative intentions. By contrast, the written judgment/rewording and oral summarization tasks were classified as moderate-Need tasks because the target words were supplied through the task input and the learners were required to process and use them in completing the task. In these conditions, Need was more externally imposed by the instructional requirements. Across all the four conditions, the learners had access to bilingual Persian–English dictionaries, thereby ensuring the presence of Search. In addition, all the tasks required the learners to select, adapt, or integrate lexical items into meaningful contexts, thereby creating opportunities for evaluation. Thus, the design was intended to isolate the role of Need level while keeping the cognitive generative demands of Search and Evaluation broadly comparable across conditions. Table 3 summarizes the involvement load profile of each task.

Table 3. Tasks of the study with their loads

Load / Tasks	Need	Search	Evaluation	Mode
Composition writing	Strong	Present	Strong	Written
Storytelling	Strong	Present	Strong	Spoken
Judgment and rewording	Moderate	Present	Strong	Written
Oral summarization	Moderate	Present	Strong	Spoken

3.5. Procedure

The experimental procedure was conducted in three phases: pre-task screening, task implementation, and post-task assessment.

In the pre-task screening phase, the Oxford Placement Test (OPT) was administered to determine the participants' general English proficiency. Only the learners whose scores fell in the intermediate range were included in the final sample. In addition, a vocabulary pretest was administered to the moderate-Need groups to ensure that the selected target words were largely unfamiliar to them. This pretest was not administered to the strong-Need groups because prior exposure to the target words could have undermined the authenticity of the strong-Need condition.

During the task implementation phase, all the participants completed their assigned tasks in a single 40-minute session. For the oral tasks, this duration encompassed a mental preparation phase followed by the actual oral production. During the training phase, students were asked to work together to conceptualize the task, to discuss the meaning and to commit the target words to memory. What was important was their being strictly forbidden to take any written notes during this period. After mentally formulating their output, they completed the task orally. In the meantime, the writing groups spent all of their forty minutes writing, checking, and re-writing their texts on paper. It should also be emphasized that all experimental conditions were carried out under exactly the same conditions, i.e. Participants were given equal time for all tasks, and the teacher did not give explicit explanations of the tasks, vocabulary input, or corrections when participants completed the tasks. All the sessions were also conducted by the same researcher to ensure that the procedures were identical. Moreover, the four groups of experimental meetings took place at different times and, apart from the fact that the groups were used in writing for both groups, the participants in the first group meeting did not meet their counterparts in the other two groups.

In the strong-Need conditions, learners were given a common thematically restricted topic and asked to produce either a written composition or an oral story. Whenever they encountered a lexical gap, they consulted the Oxford Online Dictionary and marked the relevant entry before incorporating the word into their production. This process involved both a receptive encounter through dictionary consultation and a productive encounter through contextualized use of the target vocabulary.

The strong-Need groups completed their tasks prior to the moderate-Need groups. After these sessions, the researchers identified the 20 most frequently marked words shared across the strong-Need groups. These words were subsequently incorporated into the tasks designed for the moderate-Need conditions. This sequential arrangement was adopted because the target words emerged from learners' actual lexical needs rather than being predetermined by the researchers.

In the moderate-Need conditions, the target words were explicitly provided within the task materials. The judgment and rewording group reformulated 20 sentences, each containing one underlined target word. The oral summarization group read a short text in which the same 20 target words were highlighted and then orally summarized the text for the class using those words. In both moderate-Need tasks, the learners were allowed to consult the Oxford Online Dictionary to confirm meanings during the task completion. Prior to the task implementation, the participants were instructed that all the 20 target words had to be used in completing the task; otherwise, the task would not be considered complete. All the participants complied with this requirement and incorporated the full set of the target words in their task performance. Across all the conditions, the written outputs were collected and the oral performances were audio-recorded to document the task completion. Immediately after the tasks, all the participants took the immediate vocabulary posttest. Three weeks later, a delayed posttest was administered to measure vocabulary retention.

Although the target words were made available in the task materials, vocabulary learning was not presented as the primary goal of the activity. The learners' main objective was to complete the communicative tasks, and vocabulary processing occurred as a by-product of task engagement. Therefore, the learning context can be characterized as task-induced incidental vocabulary learning.

One methodological issue is that, while the strong-Need participants did not come across the target words during the vocabulary pretest, the moderate-Need participants did. As previously mentioned, this asymmetrical design was a purposeful methodological requirement to maintain the strong-Need condition's purely self-driven nature. Although the pretest served only as a diagnostic tool with no definitions, feedback, or opportunities for active use, we recognize that the moderate-Need group may experience a testing effect simply from prior exposure. As a result, this design automatically creates a conservative test for our main hypothesis. The results would theoretically be biased against the strong-Need condition if there was any possible inflation in the moderate-Need group's scores as a result of pretest sensitization, subjecting its effectiveness to a more thorough assessment.

3.6. Data Analysis

The data were analyzed using SPSS. Descriptive statistics, including means and standard deviations, were first computed for all immediate and delayed vocabulary scores. To examine the individual and interactive effects of task modality, Need level, and testing time, a three-way mixed-design ANOVA was conducted.

Additionally, to illustrate individual learner trajectories and variations across the two testing times, individual data plots were generated alongside the group means.

Task Modality at written and spoken, and Need at strong and moderate levels were entered as between-subjects factors. Testing Time at immediate posttest and delayed posttest levels was entered as a within-subjects factor. Prior to the inferential analysis, the assumptions of normality and homogeneity of variance were examined. Normality was inspected through distributional indices and visual plots, and Levene's test was used to assess the homogeneity of variances across the groups. Because the within-subjects factor had only two levels, the assumption of sphericity was automatically satisfied. The effect sizes were reported using partial eta squared (η_p^2). The alpha level was set at .05.

IV. RESULTS

4.1 Descriptive Statistics

Table 4 presents the means (*M*) and standard deviations (*SD*) for the immediate and delayed vocabulary post-tests across the four task conditions. In the immediate post-test, the composition writing group (written, strong-Need) obtained the highest mean score (*M* = 17.88, *SD* = 1.46), followed by the judgment and rewording group (*M* = 16.00, *SD* = 1.63) and the storytelling group (*M* = 15.50, SD = 2.31). The oral summarization group recorded the lowest performance (*M* = 13.38, *SD* = 2.47).

Table 4. Descriptive statistics for immediate and delayed vocabulary post-tests by task mode and task

Task mode	Task	*n*	Immediate post-test		Delayed post-test	
			M	*SD*	*M*	*SD*
Oral	Storytelling	16	15.5	2.31	12.75	2.98
Written	Composition writing	16	17.88	1.46	16.88	1.86
Written	Judgment and rewording	16	16.00	1.63	13.44	2.80
Oral	Oral summarization	16	13.38	2.47	11.13	2.96
Total		64				

A similar pattern was observed in the delayed post-test. The composition writing group retained the highest scores (*M* = 16.88, *SD* = 1.86), while the storytelling (*M* = 12.75, *SD* = 2.98) and judgment and rewording groups (*M* = 13.44, *SD* = 2.80) showed comparable levels of retention. The oral summarization group again obtained the lowest scores (*M* = 11.13, *SD* = 2.96). Although these descriptive trends suggest an overall advantage for written task conditions, inferential interpretations depend on the interaction effects reported below.

4.2 Mixed-Design ANOVA Results

A 2 × 2 × 2 mixed-design ANOVA was conducted with Time (immediate vs. delayed) as a within-subjects factor and Task Mode (written vs. spoken) and Need (strong vs. moderate) as between-subjects factors. Because the within-subjects factor included only two levels, the assumption of sphericity was inherently met.

Regarding within-subjects effects, a significant main effect of Time was found, $F(1,60) = 275.205$, $p < .001$, $\eta_p^2 = .821$, confirming that vocabulary scores declined significantly from the immediate to the delayed post-test across all the conditions.

The Time × Need interaction was significant, $F(1,60) = 4.238$, $p = .044$, $\eta_p^2 = .066$, indicating that the rate of vocabulary attrition over time differed between strong-Need and moderate-Need conditions. Similarly, the Time × Mode of Task interaction was significant, $F(1,60) = 7.757$, $p = .007$, $\eta_p^2 = .114$ suggesting that written and spoken task conditions did not decay at the same rate across the two testing points.

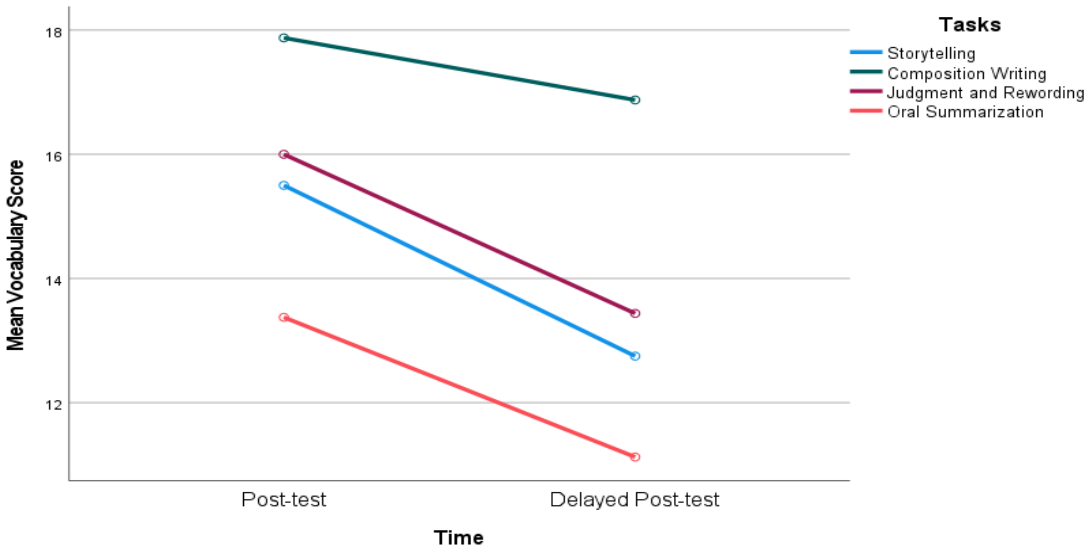


Figure 1. Mean Vocabulary Scores Across Task Conditions and Testing Times

Most notably, the three-way interaction of Time × Need × Mode of Task was significant, $F(1, 60) = 15.968, p < .001, \eta_p^2 = .211$, indicating that the pattern of retention over time differed across the four task conditions. As illustrated in Figure 3, while all the groups showed a decline in scores over time, the rate of attrition was least pronounced in the composition writing group (written, strong Need). Furthermore, Figure 4 provides a granular view of participant-level performance across the two testing points. The individual trajectory slopes demonstrate that participants in the composition writing group maintained comparatively stable scores with minimal attrition, whereas those in the judgment and rewording, storytelling, and oral summarization groups exhibited steeper downward trajectories, reflecting greater memory decay over time. To further examine this interaction, a one-way repeated-measures ANOVA with Task Type as a four-level factor confirmed a significant Time × Task Type interaction, $F(3, 60) = 9.321, p < .001, \eta_p^2 = .318$. To address potential concerns regarding the sample size ($N = 64$) and the robustness of the three-way ANOVA, a post-hoc power analysis was conducted using G*Power 3.1. The observed partial eta squared ($\eta_p^2 = 0.211$) was converted to Cohen's f effect size ($f \approx 0.517$). The analysis revealed that the statistical power ($1 - \beta$) for detecting the three-way interaction exceeded .90, which is well above the conventional threshold of 0.80. This confirms that the sample size was fully adequate to reliably detect the observed effects.

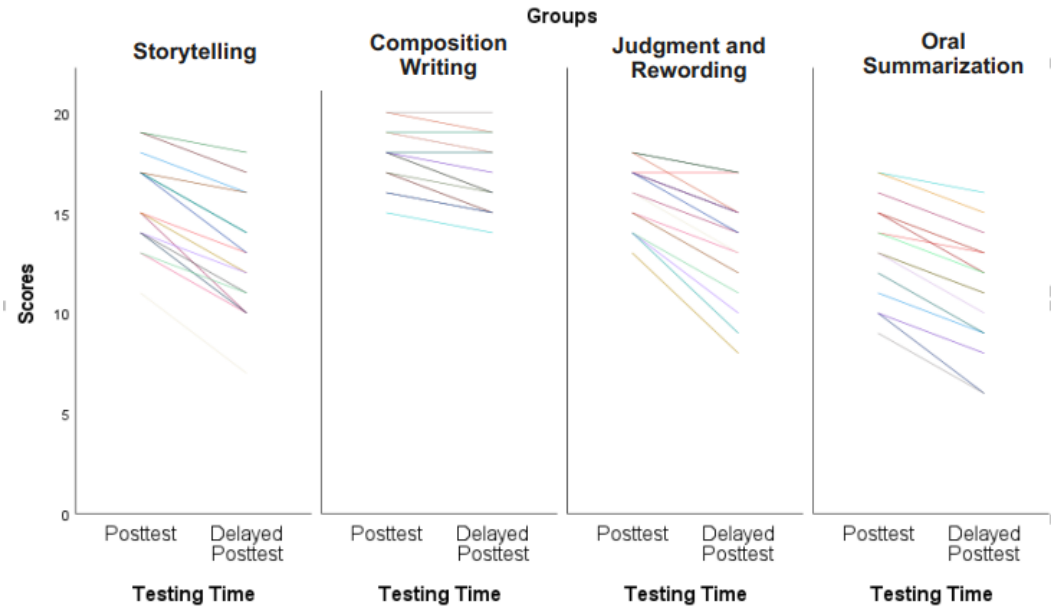


Figure 2. Individual-Level Trajectories of Test Scores Across Groups

Regarding between-subjects effects, significant main effects for both Need, $F(1,60) = 15.257$, $p < .001$, $\eta_p^2 = .203$ and Mode of Task, $F(1,60) = 24.301$, $p < .001$, $\eta_p^2 = .288$ indicating that, averaged across testing times, the strong-Need conditions yielded higher vocabulary scores than the moderate-Need conditions, and the written tasks outperformed the spoken tasks. The Need × Mode of Task interaction was not significant, $F(1,60) = .454$, $p = .503$, $\eta_p^2 = .008$ suggesting that the advantage of Need level did not differ as a function of task mode when time was not considered.

4.3. Post-hoc Comparisons

Tukey HSD post-hoc comparisons revealed that composition writing significantly outperformed all the other conditions, scoring higher than storytelling by 3.25 points ($p = .001$), higher than judgment and rewording by 2.66 points ($p = .010$), and higher than oral summarization by 5.12 points ($p < .001$). Judgment and rewording significantly outperformed oral summarization by 2.47 points ($p = .019$), but did not differ significantly from storytelling ($p = .887$). Likewise,storytellingdid not differ significantly from either oral summarization ($p = .113$) or judgment and rewording ($p = .887$). Oral summarization yielded the lowest performance overall, differing significantly from all the conditions except storytelling. These findings suggest the following hierarchy of vocabulary learning outcomes: composition writing > judgment and rewording ≈ storytelling > oral summarization.

V. DISCUSSION

The present study examined the effects of task modality and task-induced involvement on L2 vocabulary learning and retention. Overall, the findings provide qualified support for the Involvement Load Hypothesis (Laufer & Hulstijn, 2001) while also highlighting the moderating role of task modality in shaping how involvement processes contribute to vocabulary learning.

With regard to the first research question, the results indicated that the learners who performed written tasks achieved significantly greater vocabulary gains than those who completed oral tasks. This finding may be interpreted in the light of the different cognitive requirements associated with written and spoken language production. As Shin (2024) points out, oral tasks are usually performed under real-time communication pressure, which requires learners to devote significant attentional resources to maintaining fluency and conveying messages. While our data does not directly measure cognitive processes, it is plausible to hypothesize that under such conditions, fewer cognitive resources may remain available for the careful processing of unfamiliar lexical items. By contrast, written tasks provide learners with more opportunities to plan, monitor and revise their language. This slower processing rate can enable students to engage more deeply in form-related relationships and to elaborate on new lexical items in long-term memory (Gallardo del Puerto and Martínez Adrián, 2022). In the context of the current study, although the oral groups were allotted 40 minutes, this duration encompassed both mental preparation and actual performance strictly without the aid of written notes. Drawing on Levelt's (1989) speech production model, this lack of external storage heavily burdened their working memory. Participants had to simultaneously retrieve, encode, and articulate words online, a cognitive burden that ultimately took away from the depth of the vocabulary processing. Looking through the lens of Cognitive Load Theory (Sweller, 1988), the task in real time imposes a significant external cognitive load, leaving a limited working memory capacity (germane load) to process the relationships in the form.

The most important finding of the study relates to the second research question, which revealed a significant three-way interaction among testing time, task modality, and involvement-related task conditions. This interaction suggests that the durability of vocabulary learning depends not only on the degree of task-induced involvement but also on the modality through which that involvement is enacted. In particular, the composition-writing condition, combining written production with a strong Need to use target vocabulary, resulted in the most stable learning outcomes and showed the smallest decline between the immediate and delayed post-tests. Importantly, this time-dependent dynamic theoretically explains why the two-way interaction (Need × Modality) was non-significant (p= 0.503), while the three-way interaction involving Time was significant. The non-significant two-way interaction implies that, on average across both testing times, the effect of Need did not fundamentally differ by modality. However, the significant three-way interaction reveals that the relationship between Need and Modality is highly sensitive to the passage of time. In the immediate post-test, learners could likely rely on short-term

memory traces, which temporarily masked the differential effects of modality. Over time, however, the fragile encoding resulting from the cognitive overload in oral tasks decayed more rapidly. In contrast, the deeper, self-paced encoding established during written tasks demonstrated significant resilience, meaning the modifying effect of modality on involvement load only truly materialized after a delay.

More specifically, the results suggest that tasks imposing a strong Need may be especially effective when implemented through written production. In the composition-writing condition, the Need for the target words was internally generated by the learners rather than externally imposed. Instead of having to include a predetermined list of words, students write freely on the subject and encounter real lexical gaps when trying to express their intended meanings. These self-generated lexical gaps which subsequently formed the target words represent a 'strong Need' in its most authentic form. We tentatively interpret this authentic, learner-driven need as a catalyst for deeper engagement with lexical items and more deliberate processing of form–meaning relationships. However, it is crucial to emphasize that oral tasks should not be viewed merely as inferior to written tasks. Rather, they serve complementary pedagogical functions. While written tasks in our study appeared more conducive to the precise consolidation and retention of discrete vocabulary items, oral storytelling tasks are indispensable for developing real-time fluency, automaticity, and broader communicative competence. These findings offer an extension to the Involvement Load Hypothesis. While the ILH predicts that higher levels of involvement should facilitate vocabulary learning, the present results suggest that within the specific parameters of this task set, modality conditions the effect of the 'Need' component. In written tasks, learners are able to pause, consult lexical resources, and reconsider their lexical choices, which may allow the Search and Evaluation components of involvement to operate more thoroughly. By contrast, oral tasks often require rapid online processing, which may limit the depth to which these components are processed during the performance of the task. This cognitive load perspective also elucidates the contradictory finding within the oral medium itself: oral storytelling (inducing a strong, internal Need) did not significantly outperform oral summarization (inducing a moderate, external Need) in the immediate post-test. According to the core tenets of the ILH, the higher involvement index of storytelling should have resulted in better retention. However, to understand why internal need has not led to better outcomes, we need to consider the different cognitive costs of these conditions. From the perspective of Cognitive Load Theory, generating an internal Need demands vastly more cognitive resources than responding to an *external Need*. In storytelling (internal Need), learners must conceptualize original ideas from scratch and independently identify their own lexical gaps, which imposes a severe intrinsic and extraneous cognitive load. Conversely, summarization (external Need) relies on a pre-existing source text that provides a conceptual scaffold, thereby reducing the burden of message generation. Because internal Need is highly resource-intensive, combining it with the real-time pressures of oral production quickly saturated the learners' working memory. This created a cognitive 'bottleneck' (Skehan, 1998); any residual attentional resources (germane load) were entirely consumed by the mechanics of speech production rather than vocabulary processing. Consequently, the theoretical benefits of the strong, internally-driven involvement load in the storytelling task were neutralized, preventing deeper encoding of the target words.

From this perspective, the way in which the tasks are performed may be a factor in the situation which influences the way in which the learners process the meaning of the words using the depth model. Written tasks seem likely to give learners more opportunity to focus and to explore their own lexical choices and resources, which could improve the learning outcomes of students with a strong Need. As a result, the engagement load may also work, perhaps not for all types of tasks, and may vary depending on how the task requirements affect the cognitive processing capacity of the learner.

These findings are also consistent with previous research showing that written production may support a deeper lexical processing than spoken production. Laufer and Rozovski-Roitblat (2011), for example, report a larger increase in vocabulary for tasks requiring written work and attribute this advantage to the increased possibilities for lexical decision making and checking. Similarly, Zhang and Zhang (2022) found that written tasks facilitated a stronger retention of vocabulary than oral tasks. The present study extends this line of research by showing that the benefits of writing are particularly strong when the tasks involve a greater incentive to use the target vocabulary. The educational implications of these findings and the methodological limitations of this study are discussed in detail in the following concluding sections.

VI. CONCLUSION AND IMPLICATIONS

In summary, the findings of this study highlight the importance of considering both the task-induced involvement and the mode of the learning of vocabulary. While tasks that require greater involvement may facilitate the learning of vocabulary, the results suggest that the way learners perform these tasks may affect the retention of vocabulary. In particular, written tasks that require the active use of target words seem to create favorable conditions for a more profound lexical processing and a more stable learning outcome. These findings highlight the value of integrating insights from the Involvement Load Hypothesis with task-based modal perspective to better understand how learning tasks can support efficient vocabulary acquisition.

Pedagogical Implications

In terms of pedagogical implications, the findings of this study offer practical insights for vocabulary instruction. Specifically, the significant three-way interaction among Time, Need, and Mode highlights the importance of strategic task design in maximizing long-term retention. While various tasks may yield comparable immediate vocabulary gains, employing tasks with a 'strong Need' component in a written mode (e.g., essay writing or storytelling) appears to act as a cognitive anchor that significantly mitigates vocabulary attrition over time. Therefore, to ensure that newly encountered words are not quickly forgotten, instructors are encouraged to design written output tasks that inherently require learners to independently evaluate and incorporate target words to achieve their communicative goals, thereby fostering deeper and more durable lexical processing. Furthermore, since our findings suggest that the Involvement Load Hypothesis is modality-dependent, teachers must exercise caution when implementing oral tasks. Because real-time speaking inherently imposes a high cognitive load, demanding a 'strong Need' in oral production may overwhelm learners' working memory and create a cognitive bottleneck, ultimately hindering vocabulary consolidation. Consequently, when oral tasks are utilized for vocabulary practice, instructors should either opt for tasks with moderate involvement load or provide adequate pre-task planning time to alleviate extraneous cognitive burdens.

Limitations and Future Research

Several limitations should be acknowledged when interpreting these findings. First, the retention interval (three weeks) was relatively short, so the durability of the observed gains over longer periods remains unknown. The second limitation concerns the operationalization of involvement. While the tasks differed in the strength of the Need component, the Search and Evaluation components were held constant. As a result, the present design isolates the contribution of Need but does not capture how the components might interact when manipulated simultaneously. Moreover, the pretest was administered only to moderate-Need groups in order to preserve the authenticity of strong-Need task conditions. Finally, the relatively modest sample size (N = 64, approximately 16 per group) limited statistical power for detecting higher-order effects, so the three-way interaction in particular should be interpreted with caution. Additionally, learning outcomes were assessed relying solely on receptive vocabulary measures. Thus, the extent to which these involvement load conditions facilitate productive vocabulary knowledge remains unexamined.

Although the pretest was intended to be diagnostic rather than instructional, incidental exposure to the target items cannot be entirely excluded. This imbalance partially confounds the Need variable with prior exposure, and the Need comparisons should therefore be read with this caveat in mind. In addition, because intact classes rather than individually randomized participants were assigned to the experimental conditions, the quasi-experimental design constrains strong causal interpretation. Finally, the relatively small sample size (N = 64, approximately 16 per group) limited statistical power for detecting higher-order effects, so the three-way interaction in particular should be interpreted with caution.

Future research could address these limitations in several ways. Studies might examine longer-term retention, the interaction of involvement load with learner proficiency, and the simultaneous manipulation of multiple involvement components. To address the reliance on receptive tests, future studies should also incorporate productive vocabulary measures to provide a more comprehensive picture of lexical development.

Furthermore, as the present study relied on post hoc interpretations of outcome measures (i.e., vocabulary test scores), the exact cognitive mechanisms driving these results remain somewhat

speculative. Future research would benefit immensely from incorporating process-tracing methodologies. Employing techniques such as think-aloud protocols, keystroke logging, or eye-tracking would provide direct, real-time evidence of how learners cognitively navigate and process involvement load across different task modalities. Finally, examining these variables within digital or collaborative learning environments would also help clarify how task modality and involvement load operate across a broader range of instructional settings.

Use of Generative AI

During the preparation of this work, the authors used Gemini solely for the purpose of language editing, proofreading, and improving grammatical accuracy. After using this tool, the authors reviewed and edited the content as needed. They take full responsibility for the published content.

AUTHORS' BIOGRAPHIES

Mahdieh Sattar is a Ph.D. candidate in Teaching English as a Foreign Language (TEFL) at Islamic Azad University, Maybod Branch, Iran. Her primary research interests include TaskBased Language Teaching (TBLT), second language assessment, and second language vocabulary acquisition.

Mohammad Golshan is an Assistant Professor in the Department of English Language at Islamic Azad University, Maybod Branch, Iran. He holds a Ph.D. in English Language Teaching from the University of Mysore, India. His main research interests include brainbased learning, Task-Based Language Teaching (TBLT), English for Academic Purposes (EAP), and second language skills pedagogy. He has published his research in various national and international academic journals.

Ali Akbar Moeen is an Assistant Professor in the Department of English Language at Islamic Azad University, Maybod Branch, Iran. He holds a Ph.D. in English Language Teaching. His primary research interests include applied linguistics, second language acquisition, and language pedagogy.

REFERENCES

Adams, R., & Ross-Feldman, L. (2008). Does writing influence learner attention to form? In D. Belcher & A. Hirvela (Eds.), *The Oral-literate Connection: Perspectives on L2 Speaking, Writing, and Other Media Interactions* (pp. 243–266). University of Michigan Press.

Baralt, M., & Gómez, J. M. (2017). Task-based language teaching online: A guide for teachers. *Language Learning & Technology, 21*(3), 28–43. https://www.learntechlib.org/p/191544/

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*(6), 671–684. https://doi.org/10.1016/S0022-5371(72)80001-X

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford University Press.

Ellis, R. (Ed.). (2005). *Planning and task performance in a second language*. John Benjamins Publishing.

Ellis, R. (2018). *Reflections on task-based language teaching*. Multilingual Matters.

Ellis, R., Skehan, P., Li, S., Shintani, N., & Lambert, C. (2019). *Task-based language teaching: Theory and practice*. Cambridge University Press.

Gallardo-del-Puerto, F., & Martínez-Adrián, M. (2022). Task modality effects on the production and elaboration of language-related episodes: A study on schoolchildren's interactions in a foreign language. *Languages*, *7*(4), 306. https://doi.org/10.1075/resla.20020.gal

Hu, H. C. M., & Nassaji, H. (2016). Effective vocabulary learning tasks: Involvement load hypothesis versus technique feature analysis. *System, 56*, 28–39.

https://doi.org/10.1016/j.system.2015.11.003

Huang, S., Willson, V., & Eslami, Z. (2012). The effects of task involvement load on L2 incidental vocabulary learning: A meta-analytic study. *The Modern Language Journal, 96*(4), 544–566. https://doi.org/10.1111/j.1540-4781.2012.01392.x

Hulstijn, J. H. (2019). An individual-differences framework for comparing nonnative with native speakers: Perspectives from BLC theory. *Language Learning, 69*(S1), 6–27. https://doi.org/10.1111/lang.12317

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

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

Laufer, B., & Rozovski-Roitblat, B. (2011). Incidental vocabulary acquisition: The effects of task type, word occurrence, and their combination. *Language Teaching Research, 15*(4), 391–411. https://doi.org/10.1177/1362168811412019

Liu, S., & Reynolds, B. L. (2022). Empirical Support for the Involvement Load Hypothesis (ILH): A Systematic Review. *Behavioral sciences (Basel, Switzerland)*, *12*(10), 354. https://doi.org/10.3390/bs12100354

Long, M. H. (2015). *Second language acquisition and task-based language teaching*. John Wiley & Sons.

Martínez-Fernández, A. (2008). Revisiting the involvement load hypothesis: Awareness, type of task and type of item. In M. Bowles, R. Foote, S. Perpiñán, & R. Bhatt (Eds.), *Selected proceedings of the 2007 Second Language Research Forum* (pp. 210–228). Cascadilla Proceedings Project.

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

Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge University Press.

Nunan, D. (2004). *Task-based language teaching*. Cambridge University Press.

Peters, E., & Webb, S. (2018). Incidental Vocabulary Acquisition through Viewing L2 Television and Factors That Affect Learning. Studies in Second Language Acquisition, 40, 551-577. https://doi.org/10.1017/S0272263117000407

Pichette, F., de Serres, L., & Lafontaine, M. (2012). Sentence reading and writing for second language vocabulary acquisition. *Applied Linguistics, 33*(1), 66–82. https://doi.org/10.1093/applin/amr044

Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language Teaching Research, 12*(3), 329–363. https://doi.org/10.1177/1362168808089921

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Palgrave Macmillan. https://doi.org/10.1057/9780230293977

Schmitt, N. (2019). Understanding vocabulary acquisition, instruction, and assessment: A research agenda. *Language Teaching*, *52*(2), 261–274. https://doi.org/10.1017/S026144481900005X

Shin, H. (2024). Instructional task modality and teacher effects on L2 vocabulary learning:

Evidence from adolescent learners. Language Teaching Research, 28(4), 567–586.

https://doi.org/10.1177/13621688241253617

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford University Press.

Skehan, P. (2018). *Second language task-based performance: Theory, research, assessment*. Routledge.

Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidlhofer (Eds.), *Principle and practice in applied linguistics* (pp. 125–144). Oxford University Press.

Swain, M. (2005). The output hypothesis: Theory and research. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 471–483). Lawrence Erlbaum Associates.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

Un-udom, N. (2018). The effects of task-induced involvement on EFL vocabulary learning. *Chophayom Journal*, *29*(3), 181-196.

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

Webb, S. (2005). Receptive and productive vocabulary learning: The effects of reading and writing on word knowledge. *Studies in Second Language Acquisition*, *27*(1), 33–52. https://doi.org/10.1017/s027226310505002x

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

Webb, S., & Nation, I. S. P. (2017). How vocabulary is learned. Oxford, UK: Oxford

University Press.

Webb, S., Yanagisawa, A., & Uchihara, T. (2020). How effective are intentional vocabulary-learning activities? A meta-analysis. *Modern Language Journal, 104*(4), 715–738. https://doi.org/10.1111/modl.12671

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

Zou, D. (2017). Gamified vocabulary learning: Investigating the effectiveness of digital game-based learning. Educational Technology & Society, 20(1), 80–92.

Zou, D., Wang, M., Xie, H., Cheng, G., Wang, F. L., & Lee, L. (2021). A comparative study on linguistic theories for modeling EFL learners: Facilitating personalized vocabulary learning via task recommendations. *Interactive Learning Environments*, *29*(2), 270–282. https://doi.org/10.1080/10494820.2020.1789178