



The effects of task-induced involvement load and input modality on incidental vocabulary learning

Cansu Kıvrak^{a*} , Dilek Uygun Gökmen^b 

^a Yalova University, Yalova Üniversitesi Yabancı Diller Yüksekokulu, Yalova 77200, Turkey

^b Marmara University, Marmara Üniversitesi Atatürk Eğitim Fakültesi, İstanbul 34722, Turkey

APA Citation:

Kıvrak, C., & Uygun Gökmen, D. (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.

Submission Date: 29/07/2019

Acceptance Date: 13/10/2019

Abstract

Knowing enough vocabulary is crucial for second language learners to comprehend and produce the language. Therefore, a substantial bulk of research has aimed at finding the most effective ways to acquire new words. The Task-induced Involvement Load Hypothesis (Laufer & Hulstijn, 2001) states that vocabulary tasks are efficient when they induce higher learner involvement. The present study investigated how tasks with the same involvement load but different input modalities (written vs. audiovisual) affect vocabulary learning at different proficiency levels. 236 Turkish (lower- or upper-intermediate) EFL learners performed six vocabulary tasks with three involvement loads and two types of input. Four tasks included gap filling or sentence writing with eight target words after reading a text or watching a video, while two tasks involved reading or video comprehension only. Productive and receptive word knowledge was tested through vocabulary post-tests which required the target form or meaning. Data were obtained by counting the percentage/number of the correct forms and meanings. The results showed that for receptive word knowledge, sentence writing (higher involvement load) was more effective than gap filling (lower involvement load) for both levels regardless of input type, but the audiovisual input fostered more knowledge among the upper-intermediate learners. As for productive word knowledge, the pairing of gap filling with the written input and sentence writing with the audiovisual input was more effective for both levels. While these findings partially confirm the predictions for receptive word knowledge, they also highlight the role that input modality plays in productive word knowledge.

© 2019 JLLS and the Authors - Published by JLLS.

Keywords: Incidental vocabulary learning; task-induced involvement load; task efficacy; input type; productive and receptive word knowledge

1. Introduction

Vocabulary is one of the most important elements of foreign or second language acquisition because acquiring a vast number of words has a facilitating role not only in comprehension but also in production of the language. To emphasize how vocabulary knowledge and language acquisition are interconnected, Schmitt (2000) stated that “lexical knowledge is central to communicative competence and to the acquisition of a second language” (p. 55). Nation (2001) also regarded vocabulary knowledge and language use as complementary. Knowledge of words positively affects language use, and language use

* Cansu Kıvrak. Tel.: +90-226-815-5675
E-mail address: cansukivrk@gmail.com

results in an increase in vocabulary size. Similarly, Folse (2006) acknowledged that learning a great deal of vocabulary is essential for L2 learners to perform better in four major language skills. Therefore, it is clear that having a large body of vocabulary plays a crucial part in helping L2 learners express themselves more clearly and understand the messages coming from their interlocutors more easily.

Research on L2 vocabulary acquisition has revealed that L2 learners need to know at least 3,000 word families to understand spoken texts (Webb & Rodgers, 2009) and 8,000 word families to understand written texts (Nation, 2006). These two numbers only represent 95% lexical coverage for spoken texts and 98% lexical coverage for written texts. Even such large figures are not sufficient to reach 100% lexical coverage. As a result, it is obvious that learning vocabulary is one of the biggest obstacles that L2 learners have to overcome in order to be successful language users.

For L2 vocabulary learning, two central views have been proposed. The first is intentional learning which is described as “the deliberate committing to memory of thousands of words (their meaning, sound, and spelling)” (Hulstijn, 2003, p. 349). The second one is incidental learning which is defined as “picking up” words while engaging in a communicative task with a focus on meaning. Compared to intentional learning, incidental learning has been regarded as the major source of L2 vocabulary knowledge (Nation, 2013). This is mainly because native speakers also acquire most of their lexicon through daily activities in incidental settings before intentionally learning their mother tongue at school. Moreover, L2 learners need to acquire large quantities of words for communicative competence, which cannot be achieved only through intentional vocabulary activities (e.g., word substitution, crossword puzzles, etc.). That is why incidental learning receives great attention in L2 vocabulary acquisition.

Reading has been considered to be a primary resource for incidental vocabulary learning (Brown, Waring & Donkaewbua, 2008; Horst, Cobb & Meara, 1998; Huckin & Coady, 1999; Krashen 1993; Pigada & Schmitt, 2006; Waring & Takaki, 2003). However, incidental vocabulary learning through reading has a rather low pick-up rate. For example, Waring and Takaki (2003) found that only one word can be acquired from one hour of reading. Horst, Cobb and Meara (1998) further claimed that L2 learners need to encounter a new word about 8 times to be able to acquire it. Thus, it is suggested that learning words through mere reading is a rather time-consuming process.

Research has also focused on listening as a medium for incidental vocabulary acquisition (Lin, 2010; Jing & Jianbin, 2009; van Zeeland & Schmitt, 2013; Vidal, 2011). Studies comparing the benefits of listening to reading have found that reading initially leads to more word gains than listening, but vocabulary knowledge gained from listening seems to be retained better. However, Vidal (2011) stated that more encounters are required in listening (5-6 times) than reading (2-3 times) for more durable effects. Van Zeeland and Schmitt (2013) also stated that meaning is more difficult to be acquired through listening. Considering these drawbacks, it is reasonable to consult to other ways of learning vocabulary incidentally.

When reading or listening appears not to be a particularly effective approach to expanding L2 word knowledge, Paribakht and Wesche (1997) noted that L2 learners cannot be expected to learn a considerable amount of vocabulary without any instructional intervention. Thus, Laufer (2003) suggested that this instruction can be achieved through form-focused vocabulary tasks that would direct learners' attention to specifically selected lexical items. Later, Laufer and Hulstijn (2001) proposed the Task-induced Involvement Load Hypothesis to determine the superiority of certain word-focused vocabulary tasks over others in fostering L2 vocabulary knowledge. The hypothesis assumes that a vocabulary task can be efficient if it induces more learner involvement. In other words, if learners perform a vocabulary task that requires them to deal with the target words, the likelihood of retrieving these words later will be higher.

1.1. Literature review

1.1.1. The Involvement Load Hypothesis

With their depth of processing model in cognitive psychology, Craik and Lockhart (1972) aimed to explain how the degree of mental effort exhibited while attending new information can foster its long-term retention. The central idea of the model emphasizes that information is encoded in two levels (shallow or deep) and the degree of decay or retention is determined by the shallowness or depth of the cognitive processing. If new information is processed deeply during its first encounter, it is more likely to be stored because its memory trace will last longer. Craik and Tulving (1975) further expanded this framework by adding the component of elaboration or richness with which new information is initially encoded; that is, how it is enriched by a wide range of properties such as denotation, connotation and association. Without this elaborate processing, it is highly unlikely that the new information will be easily retrieved later.

Laufer and Hulstijn (2001) decided to incorporate the notions of depth of processing (Craik & Lockhart, 1972) and elaboration (Craik & Tulving, 1975) into L2 vocabulary learning through form-focused vocabulary tasks. They thought that if L2 learners elaborate on unfamiliar words (new information) more semantically (at a deeper level), their chance of recalling these words (retention rate) will increase. However, they also criticized the model because it does not clearly describe what constitutes a processing level and how learners process new words at each level. They noted that the construct of depth requires a concrete measure to be able to define and assess these levels. Therefore, they proposed the Involvement Load Hypothesis (ILH) to both solve this ambiguity and determine the cognitive load of vocabulary tasks. The hypothesis states that word retention is conditional upon the degree of cognitive processing and amount of learner involvement a task induces. If a task is engaging and learners are cognitively involved, there is a greater chance that new words learned through this task will be recalled better.

The ILH consists of three components: need, search and evaluation. Need is the non-cognitive dimension of the hypothesis. It is moderate if it is imposed by an external agent such as the teacher, but it is strong when the learner is intrinsically motivated to meet the task requirements. Search represents the need to figure out the meaning of an unknown word by looking it up in a dictionary or consulting to external sources such as the teacher or peers. It is absent when a learner is supplied with the target definitions. However, search occurs when a learner makes an attempt to discover the meaning of a word through dictionary use. Evaluation entails making the right decision to find the appropriate meaning or form of a particular word in relation to the others in a context or its further definitions in a dictionary. Moderate evaluation requires distinguishing between the multiple meanings or forms to find the best one. Strong evaluation, on the other hand, involves combining the new word with others to create an original sentence or text.

Based on their absence/presence, each component receives a score called an involvement index. This index can be measured through such points or symbols as 0 or (-) (i.e., the component is absent), 1 or (+) (i.e., it is moderately present) and 2 or (++) (i.e., it is strongly present). The involvement index or load of a task is determined by the sum of these scores (see Table 1 for sample tasks). For example, a reading comprehension task with the target words glossed in the margin gets an index of 1 because need is moderate (1) (induced by the text), but search and evaluation are absent (0) (as learners are not asked to find or negotiate the meaning of the target words). However, a sentence-writing task with the glossed words receives an index of 3 because need is moderate (1) (induced by the teacher), search is absent (0) (since the definitions of the target words are given), but evaluation is strong (2) (as learners are required to use the target words to produce novel sentences). By grading vocabulary tasks based on their involvement loads, Laufer and Hulstijn (2001) aimed to evaluate the efficacy of a task in terms of its

contribution to the retention of new words in an incidental setting. They claimed that tasks with higher involvement indexes are the most effective ones yielding greater vocabulary retention.

Table 1. Task-induced involvement indexes for various tasks

Sample tasks	Need	Search	Evaluation	Index
1. A reading comprehension with glossed words but not necessary for the task	0	0	0	0
2. A reading comprehension with glossed words required for the task	1	0	0	1
3. A reading comprehension with fill-in	1	0	1	2
4. Writing sentences with a list of glossed words	1	0	2	3
5. Writing a composition with a list of the words (L1 translations are given only.)	1	1	2	4

Source. Laufer and Hulstijn (2001)

1.1.2. Empirical evidence for the ILH

To present empirical support for the ILH, Hulstijn and Laufer (2001) conducted two parallel studies with 186 Dutch and Hebrew advanced EFL learners. They implemented three tasks in each setting. Task 1 involved a reading comprehension with 10 target words/phrases glossed in L1 in the margin. The group assigned to this task read a text and answered 10 comprehension questions. Their task induced moderate need (1) but no search (0) or evaluation (0), so the index of the task was 1. Task 2 (a reading comprehension plus gap-filling) necessitated reading the same passage with the target words omitted and filling them in with the help of the glosses. The task induced moderate need (1) and evaluation (1) but no search (0). Its involvement index was 2. The group assigned to Task 3 wrote a composition incorporating the target words whose definitions were provided, so the task had moderate need (1), no search (0) but strong evaluation (2). Its involvement load was 3. On the immediate and delayed post-tests, the learners wrote an L1 equivalent/L2 explanation for the target items. The results indicated that the composition group (Task 3) significantly outperformed the others on both post-tests in both studies. Although there was no difference between Task 1 and Task 2 in the Dutch setting, the learners of Task 2 significantly outscored those of Task 1 on both post-tests in the Hebrew experiment. As a result, the study provided strong support for the hypothesis because the composition-writing task with the highest involvement index fostered more vocabulary learning.

With 79 undergraduate beginning learners of Spanish as L2, Keating (2008) replicated Hulstijn and Laufer's (2001) study to explore whether their predictions can be validated both in active and passive word knowledge for low-proficiency learners. Three tasks were implemented with eight Spanish pseudowords. They were a glossed reading comprehension (Task 1), a reading comprehension plus gap-filling (Task 2) and writing sentences with the target words (Task 3). The tasks had an index of 1, 2 and 3 respectively. The participants translated eight English sentences containing the target words into Spanish on the active recall post-tests, while they supplied L1 correspondents on the passive recall post-tests. Data analysis revealed that partially in line with the hypothesis, Task 2 and Task 3 promoted more passive word knowledge than Task 1, but Task 3 was not most productive. Fully verifying the hypothesis, the scores of immediate active recall were highest in Task 3 followed by Task 2 and Task 1. However, Task 3 could not maintain its superiority on the delayed post-test as Task 2 received the highest scores. To explain this finding, Keating (2008) stated that it is no surprise to observe a decrease in knowledge and sentence writing may not trigger the kind of deep processing that occurs in composition writing. Nonetheless, it was concluded that the hypothesis can be generalized to less proficient L2 learners as well.

Kim (2011) conducted two experiments with 104 ESL learners to test the hypothesis across various language levels. The participants were divided into two groups: less proficient (Intensive English

Program) and more proficient (Undergraduate). In the first experiment, three tasks (a reading comprehension with graphic organizers (1), a reading comprehension plus fill-in (2) and writing a composition (3)) were compared. For the second experiment, two tasks (sentence writing and composition writing) inducing the same involvement load (3) were compared to verify Laufer and Hulstijn's (2001) claim that tasks sharing an identical index are equally effective in vocabulary retention. The data collected through Paribakht and Wesche's (1993) Vocabulary Knowledge Scale showed that on the immediate and delayed post-tests of the first experiment, the composition groups in both levels scored significantly higher than the gap-fill groups which in turn scored higher than the reading comprehension groups. Thus, Kim (2011) noted that strong evaluation may have a far greater influence in the overall task load although Hulstijn and Laufer (2001) considered the three components to be identical in weight. As for the second experiment, the composition and sentence writing groups in both levels obtained very similar scores. This finding was also in harmony with the hypothesis since the two equally-loaded tasks yielded similar word gains.

Most of the research on the ILH has centred on the effect of the evaluation component on the efficacy of reading-based tasks. However, several studies have also probed into the link between input modality and task-induced involvement load (Fatalaki, 2014; Hassanzadeh, 2016; Jing and Jianbin, 2009; Maleki, 2012). In one of these studies, Jing and Jianbin (2009) investigated whether listening comprehension tasks loaded with different degrees of involvement result in better vocabulary learning. To a group of 86 Chinese non-English major EFL learners, they applied three tasks (a listening comprehension with glosses irrelevant to the task (0), a listening comprehension with glosses relevant to the task (1) and a listening comprehension plus composition writing (3)). Two listening passages and 15 target words were selected. The results indicated that tasks with greater indexes lead to more word gains as the composition-writing group outperformed the others. In a more comprehensive study with 80 Iranian pre-intermediate English majors, Maleki (2012) implemented similar tasks involving 12 listening passages and 40 target words and reached the same conclusion.

In another study with 21 Iranian undergraduate EFL students, Hassanzadeh (2016) found that tasks designed with the audiovisual input are more effective than reading comprehension tasks. In the study, three tasks (Task A = a comprehension task with glosses (1), Task B = a fill-in task (2) and Task C = a sentence writing task (3)) were carried out through two types of input. The written input was thematically related reading texts, while the audiovisual input was topic-related TV talk shows. The learners completed the tasks through both modes. The results indicated that active word recall was significantly influenced by input mode, and the average of overall word gains from the three tasks was higher in the audiovisual input than in the written input. When the tasks were compared within each input presentation, the sentence writing task was found to be most effective followed by the gap-filling task and the comprehension task in the written condition. In the audiovisual mode, the sentence writing task yielded the greatest word recall followed by the comprehension task and the gap-filling task. These findings provided strong evidence for the superiority of videos over reading texts when such materials are used for task design.

1.2. Research questions

Following the assumptions of the ILH, the present study aimed to investigate the effect of input type (written vs. audiovisual) on the efficacy of particular vocabulary tasks with various or identical degrees of task-induced involvement load by measuring both productive and receptive word knowledge across two levels of proficiency. Its significance lies in exploring the influence of input modality on incidental vocabulary learning across different levels of L2 proficiency. The following research questions guided the study:

(1) Do vocabulary tasks with higher levels of involvement load result in more vocabulary gain than tasks with lower levels of involvement load at lower-intermediate level in terms of productive and receptive word knowledge?

(2) Do vocabulary tasks with different input types (audiovisual vs. written) yield differences in vocabulary gain at lower-intermediate level in terms of productive and receptive word knowledge?

(3) Do vocabulary tasks with higher levels of involvement load result in more vocabulary gain than tasks with lower levels of involvement load at upper-intermediate level in terms of productive and receptive word knowledge?

(4) Do vocabulary tasks with different input types (audiovisual vs. written) yield differences in vocabulary gain at upper-intermediate level in terms of productive and receptive word knowledge?

2. Method

2.1. Research design

The present study had a quasi-experimental design with eight experimental and four control groups through the implementation of twelve vocabulary tasks across two levels of language proficiency during one session and administration of vocabulary pre- and post-tests at the beginning, immediately after the treatment and two weeks later.

2.2. Participants & Setting

The participants consisted of 236 (116 male and 120 female) Turkish undergraduate EFL learners enrolled in a mandatory Intensive English Program at a preparatory school of a state university in Turkey. They were young adults ranging from 18 to 22 in age. They had been attending English classes at the institution for 5 months to be proficient enough for their departmental courses. The participants' majors included Engineering, Business Administration and International Relations.

The study involved twelve intact classes. Convenience sampling was used while selecting these classes to form the experimental and control groups. Six classes ($n = 123$) were at lower-intermediate level, while the other six ($n = 113$) were upper-intermediate (see Table 2 for the distribution). The participants were considered to be at these levels based on their scores on a placement test administered at the beginning of the year and the subsequent institutional examinations. All of the participants volunteered to be in the study, but they were not informed about its purpose or the upcoming vocabulary post-tests not to trigger any memorization of the target words.

Table 2. Distribution of the participants

Task type	Lower-intermediate	Upper-intermediate
Reading-only	21	19
Watching-only	23	20
Fill-in-by-reading	20	17
Fill-in-by-watching	20	21
Sentence-writing-by-reading	21	19
Sentence-writing-by-watching	18	17
Total	123	113

2.3. Instrument(s)

Several instruments were used in the study for research purposes. First of all, two different video podcasts as audiovisual materials and two separate reading passages as written texts were chosen for the study. A total of 16 vocabulary items were selected as target words to be used in the vocabulary tasks. Finally, vocabulary pre- and post-tests were implemented to gather data for the study.

2.3.1. Video podcasts & Reading passages

Two different video podcasts were selected as the audiovisual input for the study. Both were chosen from a well-known skills-integrated language coursebook series in Turkey. The video assigned to the lower-intermediate participants was about a county in England named Cornwall and a volunteer organization that is situated there. It lasted around 10 minutes and consisted of 1170 words. The other video that the upper-intermediate participants watched was about the city of Vancouver and its action plan to be the greenest city in the world. Its duration was 8 minutes. A total of 1055 words were included in the video.

Two different reading passages were used as the written input for the study. Both were the transcribed versions of the video podcasts, but they were adapted to look like a reading text by excluding some colloquial expressions. The reading passage assigned to the lower-intermediate participants was named ‘The Working Week’. It was an expository text about a volunteer organization. It had a total of 792 words. The upper-intermediate learners were also assigned to an expository text entitled ‘The Best City in the World’. It consisted of a total of 600 words.

2.3.2. Target words

16 target words (8 per level) were selected in total for investigation. To choose the target items, the words in the videos that were expected to be unfamiliar to the participants at both levels were identified by the researcher and tested on the vocabulary pre-tests. Based on the pre-test results, the target words ultimately assigned to the lower-intermediate group included three nouns, three adjectives and two verbs: *pager*, *current*, *pupil*, *dedicated*, *stunning*, *treacherous*, *launch* and *patrol*. As for the upper-intermediate learners, the target words consisted of the following nouns, adjectives and verbs: *display*, *creek*, *abundance*, *complacent*, *ultimate*, *conscientious*, *boast* and *implement*.

2.3.3. Vocabulary tasks

Six classes from each level were assigned to one of the six vocabulary tasks: a reading-only (RO) task, a watching-only (WO) task, a fill-in-by-reading (FR) task, a fill-in-by-watching (FW) task, a sentence-writing-by-reading (SR) task and a sentence-writing-by-watching (SW) task.

During the RO task, the participants first had a short in-class discussion with a general question based on the relevant topic and then read a reading passage with several comprehension questions to be completed afterwards. Later, they were instructed to do a post-reading activity by briefly reporting their ideas on another discussion question to their peers. Overall, it took around 35 minutes to perform the entire task. The task induced an index of 0 with no need (0), search (0) or evaluation (0) because the target words were not bolded or highlighted not to draw any attention to them. They were not relevant to the comprehension questions either, which means that the knowledge of these words was not necessary to find the correct answers.

The participants assigned to the WO task were not exposed to any written texts. Instead, they watched a video podcast that presented the same content as that of the corresponding reading passage. The subtitles were turned off so that the learners would only have access to the visual or aural cues. First, they were asked to do the same warm-up activity as the RO group. Then they were given 10 minutes to look at the comprehension questions before the first viewing. The video was played twice. After the

second viewing, the answers to the questions were checked in class and the final discussion was carried out to wrap up the task. Overall, it took approximately 40 minutes to finish the entire task. The involvement index of the task was calculated as 0 since it induced no need (0), search (0) or evaluation (0) and the learners did not need to acquire the meaning of the target words to be able to perform their task.

The participants of the FR task were provided with the same reading passage and comprehension questions, but they also received a glossary of the target words along with their Turkish equivalents and English definitions because their texts were gapped. After a brief discussion, these learners were asked to fill in their gapped reading passages with the target words. Then they checked their answers in class. This main task was followed by the completion of the comprehension questions and the final discussion. The entire task lasted for around 40 minutes and had an involvement index of 2 because need (1) was moderate, search (0) was absent and evaluation (1) was moderate. In other words, the participants were expected to tackle with the target words and check their appropriateness, but they did not have to look up any targeted meanings.

Following the warm-up discussion, the FW group was instructed to look at the provided glossary of the target words for a few minutes. Then they had to fill in these words into the blanks left out in the sentences directly taken from the relevant video podcast during the first viewing. After that, the learners checked their answers in class. The second viewing was carried out to allow the learners to answer the comprehension questions. Following the comprehension check, the learners performed the final discussion. The entire task ended in about 45 minutes. The involvement index of the task was 2 as need (1) and evaluation (1) were moderate, but there was no search (0). The target definitions were explicitly presented and the participants only needed to find out the correct sentence each target word fits in.

The SR task was made up of two parts. The first part was similar to the RO task. The participants were instructed to do the warm-up activity, read the passage and answer the comprehension questions. This part was added to provide a meaningful context for the target words and balance the variable of time-on-task. The second part involved writing sentences with the target words presented with their Turkish equivalents and English definitions. On their own, the participants were asked to form eight original sentences that would somehow reveal the meaning of the target items. When finished, several learners were told to read their sentences out loud so that the meaning of each target word would be covered at least once. Following this activity, the final discussion was held. On average, it took 50 minutes to complete the entire task. The SR task induced an index of 3 because need (1) was moderate, search (0) was absent, but evaluation (2) was strong. The participants had to compare the target words with others to create a new context. This requires a stronger involvement as opposed to comparing the target words across one another in the fill-in tasks.

The SW task also consisted of two parts. The first part resembled that of the WO task. The participants were provided with the glossary of the target words and asked to go over the L1/L2 definitions before watching their video podcast. This part ended after the learners watched their video podcast twice, answered the comprehension questions and checked them in class. The second part required the learners to create eight original sentences that incorporated the target words by indicating their meanings to some extent. Then some of these sentences were read out loud and the final discussion was held. The whole task ended in about 55 minutes. Since the SW task induced moderate need (1), no search (0) but strong evaluation (2), its involvement index was calculated as 3.

2.3.4. Vocabulary tests

A vocabulary pre-test was created for each level of proficiency to choose the eight target words and check whether the participants knew these words prior to the study. Each test involved a set of 20 words including the target words to be selected for that level and 12 distracters. In these pre-tests, the

participants were asked to provide Turkish equivalents/English definitions for the words in the list. They were told to put an “X” next to the words that they were not familiar with. The target words assigned to each level were selected based on the results of the pre-tests.

Two types of vocabulary post-tests were administered immediately after task implementation and two weeks later. The productive post-test consisted of the definitions of the eight target words delivered in both Turkish and English in an alphabetical order and required the participants to write the target words next to their correct explanations. The test shared the same definitions as the ones provided in the glossary list. The participants were instructed not to employ any guessing strategies. Instead, they were told to leave out the words that they could not recall. However, they were asked to write their partial answers even if their answers contained only a few letters. The rationale behind this instruction was that the present study aimed to measure both partial and full word gains in form. On the delayed productive post-test, the order of the definitions was randomly changed to prevent the participants from taking advantage of any recency effect.

The receptive post-test included the eight target words presented in an alphabetical order and expected the participants to write the correct L1 equivalents/L2 definitions next to the words. The participants were asked to provide clear explanations illustrating the meaning of the words. The delayed receptive post-test consisted of the target words arranged in a different order to prevent any recency effect.

2.3.5. Scoring

The scores of the productive post-test were calculated using a lexical production scoring protocol proposed by Barcroft (2000). According to this tool, points were given based on the complete and partial word knowledge by counting the number of letters correct or the number of letters present and calculating their percentages. A correct letter referred to any letter of a word written in its correct place, while a present letter referred to any letter of a word written but not in its correct place. For example, for the target word ‘current’, the following answers would get a score of .25: ‘c...’ or ‘...t’ (since at least one letter (1 of 7 = 14%) was in its correct place) and ‘urad’ (since at least 2 letters (2 of 7 = 28%) were present but not in their correct places). A full point (1) would be given to the correct form ‘current’ written next to its correct definition, but words not written or written next to an incorrect definition would receive a score of zero. If a learner wrote the same target word more than once, all answers were regarded as incorrect because the learner applied a guessing strategy. The maximum score that could be reached was 8 points.

The scores on the receptive post-test were calculated by counting the correct explanations of the target words. A full point (1) was given to a correct explanation written in L1/L2, while half a point (0.5) was given to a semantically acceptable explanation. A score of zero was given to the explanations not written or written next to an incorrect word. If a participant wrote the same definition more than once, all answers were accepted as incorrect because the participant used a guessing strategy. The maximum score that could be achieved on this post-test was 8 points.

2.4. Data collection procedures

Prior to the study, the researcher informed the other five instructors who were teaching the task groups at the time about the aim of the study. Their consent was taken. However, the participants were not told in advance that they were participating in a research study. The reason why these learners were not provided with any information regarding the purpose of the study or the upcoming vocabulary post-tests was not to encourage them to commit the target words into their memory.

In Week 1, the vocabulary pre-test was administered to all participants to test their familiarity with the target vocabulary. Based on the results, the target words were selected from the ones that were unknown to the participants. During Week 2, the researcher informed the other five instructors about how to implement the vocabulary tasks and tests assigned specifically to their classrooms. These instructors were randomly selected because the institution placed each one of them randomly in those classrooms. The six participant instructors, including the researcher, all received enough copies of the vocabulary tasks and tests in paper for their classrooms beforehand.

At the beginning of Week 3, all the participants simultaneously completed one of the six vocabulary tasks particularly assigned to their class in their regular classroom hours. Each participant completed his or her task individually in the presence of the instructor. Then all the printed materials were collected. Immediately after task completion, the participants unexpectedly took the productive and receptive post-tests in that order. If the receptive post-test had been administered first, the participants could have seen how the target forms were written. At the end of Week 5, the unannounced delayed productive and receptive post-tests were administered. The implementation procedure remained the same for both proficiency levels.

2.5. Data analysis

The data obtained from the vocabulary post-tests were analyzed by means of quantitative methods. The participants' scores were compared by using the Statistical Package for the Social Sciences (SPSS) 23.0 within each level of proficiency. To determine whether parametric tests were appropriate, all variables were tested for their normality of distribution by using the Shapiro-Wilk test. The significance level was set at $p = .05$. Since the data were not normally distributed based on the result of the Shapiro-Wilk test, the non-parametric Kruskal-Wallis test was used to determine if there were statistically significant differences between the task groups in their immediate and delayed productive and receptive vocabulary post-test scores.

3. Results

3.1. Vocabulary gain from tasks with low vs. high involvement load at lower-intermediate level

The first research question aimed to investigate whether vocabulary tasks with higher levels of involvement load result in more vocabulary gain than tasks with lower levels of involvement load at lower-intermediate level in terms of productive and receptive word knowledge. In order to find out whether the lower-intermediate learners differed in their immediate productive word gain, a Kruskal Wallis test was conducted to compare their scores on the immediate productive vocabulary post-test. The results were shown in Table 3.

Table 3. Initial gain in productive word knowledge at lower-intermediate level

	Groups	<i>N</i>	\tilde{x}	X^2	<i>p</i>
Immediate productive post-test	SR	21	5.5	76.649	< 0.001*
	SW	18	4		
	FR	20	5.38		
	FW	20	2		
	RO	21	0.75		
	WO	23	0		

Note. SR: Sentence-writing-by-reading (index = 3), SW: Sentence-writing-by-watching (index = 3), FR: Fill-in-by-reading (index = 2), FW: Fill-in-by-watching (index = 2), RO: Reading-only (index = 0), WO: Watching-only (index = 0), * $p < .05$. Scores were given out of 8 points.

Table 3 demonstrates that there was a statistically significant difference among the groups ($p < 0.001$). The RO and WO groups obtained significantly lower scores than the other groups, but no significant differences were found among the experimental groups.

To see whether the lower-intermediate learners were able to retrieve the target forms after a two-week interval, a Kruskal Wallis test was carried out to compare the scores of the task groups on the delayed productive vocabulary post-test. Table 4 presents the results.

Table 4. Retention in productive word knowledge at lower-intermediate level

	Groups	<i>N</i>	\tilde{x}	X^2	<i>p</i>
Delayed productive post-test	SR	21	2	55.726	< 0.001*
	SW	18	3.38		
	FR	20	3		
	FW	20	0		
	RO	21	0.25		
	WO	23	1		

Note. SR: Sentence-writing-by-reading (index = 3), SW: Sentence-writing-by-watching (index = 3), FR: Fill-in-by-reading (index = 2), FW: Fill-in-by-watching (index = 2), RO: Reading-only (index = 0), WO: Watching-only (index = 0), * $p < .05$. Scores were given out of 8 points.

The results indicated a statistically significant difference among the groups ($p < 0.001$). On the second measurement, the SW group significantly outscored the FW group and the control groups as well as performing slightly better than the FR group. While the vocabulary gain of the SW and FR groups demonstrated significant differences with the FW and control groups, the SR group also significantly outperformed the RO and FW groups. No other significant differences were found.

To determine whether the lower-intermediate participants acquired the eight target words, a Kruskal Wallis test was conducted. The findings were shown in Table 5.

Table 5. Initial gain in receptive word knowledge at lower-intermediate level

	Groups	<i>N</i>	\tilde{x}	X^2	<i>p</i>
Immediate receptive post-test	SR	21	7	58.164	< 0.001*
	SW	18	7		
	FR	20	6.25		
	FW	20	3.75		
	RO	21	2		
	WO	23	4		

Note. SR: Sentence-writing-by-reading (index = 3), SW: Sentence-writing-by-watching (index = 3), FR: Fill-in-by-reading (index = 2), FW: Fill-in-by-watching (index = 2), RO: Reading-only (index = 0), WO: Watching-only (index = 0), * $p < .05$. Scores were given out of 8 points.

When the immediate receptive vocabulary post-test scores were compared, it was found that the learners significantly differed in their initial gain of the correct meanings ($p < 0.001$). Both sentence-writing groups significantly outscored the FW and control groups. The FR group also significantly outperformed the control groups.

The results of another Kruskal Wallis test indicated the extent to which the lower-intermediate learners were able to recall the target words two weeks later. Table 6 shows the learners' receptive vocabulary retention on the second measurement.

Table 6. Retention in receptive word knowledge at lower-intermediate level

	Groups	<i>N</i>	\tilde{x}	X^2	<i>p</i>
Delayed receptive post-test	SR	21	7	57.753	< 0.001*
	SW	18	7.5		
	FR	20	6.25		
	FW	20	2		
	RO	21	2		

WO	23	3
----	----	---

Note. SR: Sentence-writing-by-reading (index = 3), SW: Sentence-writing-by-watching (index = 3), FR: Fill-in-by-reading (index = 2), FW: Fill-in-by-watching (index = 2), RO: Reading-only (index = 0), WO: Watching-only (index = 0), * $p < .05$. Scores were given out of 8 points.

As the results indicated, there was a statistically significant difference among the groups ($p < 0.001$). The SW, SR and FR groups retrieved more target words by significantly outperforming the remaining groups. In other words, the FW, RO and WO groups scored significantly lower on the delayed receptive vocabulary post-test.

3.2. Vocabulary gain from tasks with written vs. audiovisual input at lower-intermediate level

The second research question investigated whether vocabulary tasks with different input types (audiovisual vs. written) yield differences in vocabulary gain at lower-intermediate level in terms of productive and receptive word knowledge. The results of the Kruskal Wallis tests showed that the written input led to more productive word knowledge when it was presented through a gap filling task (index = 2), while the audiovisual input led to more productive word knowledge when it was presented through a sentence writing task (index = 3). As for receptive word knowledge, the lower-intermediate learners recalled a greater number of target words when they were exposed to the written input through a gap filling task (index = 2), whereas input type played a relatively small role with sentence writing tasks (index = 3).

3.3. Vocabulary gain from tasks with low vs. high involvement load at upper-intermediate level

The third research question aimed to explore whether vocabulary tasks with higher levels of involvement load result in more vocabulary gain than tasks with lower levels of involvement load at upper-intermediate level in terms of productive and receptive word knowledge. In order to determine whether the upper-intermediate learners differed in their immediate productive word gain, a Kruskal Wallis test was used to compare their median scores on the immediate productive vocabulary post-test. Table 7 shows this comparison.

Table 7. Initial gain in productive word knowledge at upper-intermediate level

	Groups	N	\tilde{x}	X^2	p
Immediate productive post-test	SR	19	3.75	66.841	< 0.001*
	SW	17	4.75		
	FR	17	3		
	FW	21	4.75		
	RO	19	0		
	WO	20	0		

Note. SR: Sentence-writing-by-reading (index = 3), SW: Sentence-writing-by-watching (index = 3), FR: Fill-in-by-reading (index = 2), FW: Fill-in-by-watching (index = 2), RO: Reading-only (index = 0), WO: Watching-only (index = 0), * $p < .05$. Scores were given out of 8 points.

Table 7 demonstrates that there was a statistically significant difference among the groups in their immediate productive vocabulary post-test scores ($p < 0.001$). It was seen that the control groups significantly obtained lower scores than the remaining groups. However, the experimental groups did not significantly differ in their immediate productive word gain.

To find out whether the upper-intermediate learners were able to recall the eight target forms two weeks after the implementation, their scores on the delayed productive vocabulary post-test were analyzed using a Kruskal Wallis test. Table 8 presents the results.

Table 8. Retention in productive word knowledge at upper-intermediate level

	Groups	N	\tilde{x}	X^2	p
Delayed productive post-test	SR	19	2	23.765	< 0.001*
	SW	17	2.75		
	FR	17	1.75		

FW	21	1.5
RO	19	1
WO	20	1

Note. SR: Sentence-writing-by-reading (index = 3), SW: Sentence-writing-by-watching (index = 3), FR: Fill-in-by-reading (index = 2), FW: Fill-in-by-watching (index = 2), RO: Reading-only (index = 0), WO: Watching-only (index = 0), * $p < .05$. Scores were given out of 8 points.

The results revealed a statistically significant difference among the groups ($p < 0.001$). When productive word knowledge was measured again, it was found that the SW group slightly outscored all the other groups, but their performance turned out to be significant only compared to the scores of the control groups.

Another Kruskal Wallis test was conducted to see whether there were any significant differences among the upper-intermediate learners in their immediate receptive vocabulary post-test scores. The results were provided in Table 9.

Table 9. Initial gain in receptive word knowledge at upper-intermediate level

	Groups	N	\tilde{x}	X^2	p
Immediate receptive post-test	SR	19	6	66.855	< 0.001*
	SW	17	7		
	FR	17	4		
	FW	21	6		
	RO	19	1		
	WO	20	2		

Note. SR: Sentence-writing-by-reading (index = 3), SW: Sentence-writing-by-watching (index = 3), FR: Fill-in-by-reading (index = 2), FW: Fill-in-by-watching (index = 2), RO: Reading-only (index = 0), WO: Watching-only (index = 0), * $p < .05$. Scores were given out of 8 points.

It can be deduced from Table 9 that all the groups, regardless of performing different tasks, acquired at least a few of the target words. When the immediate receptive vocabulary post-test scores were analyzed, it was found that the groups significantly differed in their initial gain of the correct meanings ($p < 0.001$). The highest median score was achieved by the SW group who significantly outperformed the learners in the control groups. Following these learners, the SR and FW groups also significantly did better than the control groups. The last significant difference was detected between the FR and RO groups.

The results of the delayed receptive vocabulary post-test showed the extent to which the upper-intermediate learners retrieved the eight target words within two weeks' time. Table 10 presents their receptive word retention on the second measurement.

Table 10. Retention in receptive word knowledge at upper-intermediate level

	Groups	N	\tilde{x}	X^2	p
Delayed receptive post-test	SR	19	5	31.863	< 0.001*
	SW	17	5		
	FR	17	3		
	FW	21	3.5		
	RO	19	2		
	WO	20	2.5		

Note. SR: Sentence-writing-by-reading (index = 3), SW: Sentence-writing-by-watching (index = 3), FR: Fill-in-by-reading (index = 2), FW: Fill-in-by-watching (index = 2), RO: Reading-only (index = 0), WO: Watching-only (index = 0), * $p < .05$. Scores were given out of 8 points.

The results indicated that there was a statistically significant difference among the groups ($p < 0.001$). The SW and SR groups both achieved significantly better than the control groups. However, the only significant difference among the experimental groups was between the SR and FR groups.

3.4. Vocabulary gain from tasks with written vs. audiovisual input at upper-intermediate level

The last research question investigated whether vocabulary tasks with different input modalities (audiovisual vs. written) yield differences in vocabulary gain at upper-intermediate level in terms of productive and receptive word knowledge. The results of the Kruskal Wallis tests showed that the

learners who completed their tasks through the medium of watching (SW and FW) attained greater initial productive word knowledge than those who performed the same tasks through the medium of reading (SR and FR). However, the audiovisual input led to slightly more retention of form with the sentence writing tasks, while being exposed to the written input during the gap filling tasks enhanced more productive word knowledge over time. As for receptive word knowledge, all the task groups initially acquired more target words through the audiovisual input than the written input. However, over time the audiovisual input fostered more vocabulary retention than the written input through the gap filling tasks, while no input difference was found with the sentence writing tasks.

4. Discussion

4.1. Vocabulary gain from variously loaded tasks at lower-intermediate level

The first research question of the study aimed to explore whether vocabulary tasks with higher levels of involvement load result in more vocabulary gain than tasks with lower levels of involvement load at lower-intermediate level in terms of productive and receptive word knowledge. For productive word knowledge, the results indicated that sentence writing tasks (index = 3) yield more initial knowledge than fill-in tasks (index = 2) only through the same input. However, both tasks enhance more productive word knowledge than reading/watching-only tasks (index = 0). For the retention of form, the superiority of sentence writing is still maintained through the audiovisual input. This means that productive word knowledge is fostered more when L2 learners write original sentences with the target words. This is because the attempt to write semantically appropriate sentences might encourage learners to acquire them. The strong degree of the evaluation component induced by sentence writing tasks might be another factor for their efficacy. As Kim (2011) claimed, the three components of the ILH may not be equal in value. Strong evaluation can contribute to vocabulary enhancement more because it generates more cognitive effort and thus facilitates retention.

Gap filling tasks can also help lower-proficiency learners attain a good amount of productive word knowledge as revealed in this study with the efficacy of the FR task. Several justifications can be made for this finding. First, the learners in this study were accustomed to doing fill-in-the-blank exercises due to the instructional approach of their course book, so they might be inclined to commit the target forms into memory during gapped reading. The focus on writing the suitable target words during the task might have also prompted these learners to acquire the meaning as well. Moreover, with less proficient L2 learners more cognitive resources are needed for task completion (Boers, Eyckmans & Godfroid, 2007). However, such learners employ fewer cognitive strategies and less effectively (Anderson, 1991; Green & Oxford, 1995). This relationship between L2 proficiency and cognitive ability can also explain why the FR task (less cognitively challenging) was as effective as the sentence writing tasks (more cognitively challenging) at lower-intermediate level.

On the other hand, the FW task did not create the same effect as the FR task did. Contrary to the predictions of the ILH, the FW group scored lower than almost all the other groups on both productive and receptive vocabulary post-tests. This is striking because these learners were supposed to outperform the control groups at least owing to their higher involvement index. This unexpected finding can be attributed to their poor listening skills. These learners struggled while keeping track of the speech flow in their video podcast, so they were unable to identify the target words most of the time. They lacked the necessary bottom-up skills to decode individual words. As a result, they could not supply most of the correct forms and meanings on the post-tests.

With respect to receptive vocabulary gain and retention, it was found that both sentence writing tasks (index = 3) yielded a greater amount of recall than gap filling tasks (index = 2). This finding is consistent

with those of the earlier studies conducted by Hulstijn and Laufer (2001) and Kim (2011). They also concluded that the more involving task (sentence writing) is more effective than the less involving task (gap filling). As Kim (2011) suggested, although involving tasks may be cognitively challenging for less proficient L2 learners due to their limited language abilities, they can still be favoured by these learners. Similarly, Swain's (1985) Output Hypothesis stated that the act of producing makes L2 learners become aware of the gap in their linguistic knowledge and try to compensate for it. Thus, it is obvious why the sentence writing groups reflected on the target definitions more. They wanted to convey an accurate and meaningful message with their sentences. This helped them recall the target words with less difficulty than the gap filling groups.

When reading-only and watching-only tasks are considered, it can be inferred that such tasks play a relatively insignificant role in facilitating the learning of form and meaning. Fully verifying the ILH, these tasks resulted in a rather low amount of productive word knowledge and a slightly moderate amount of receptive word knowledge. This is because these learners were not directly exposed to the target words by means of any enhancing effects such as highlighting or glossing, so they did not notice the target items. As Schmidt (1990) suggested, L2 learners need to notice the new information to learn it. Likewise, Laufer (2003) stated that explicit focus is indispensable for form-meaning associations. This explains why the control groups did not acquire most of the target words.

Partially confirming the assumptions of the ILH, the findings of the lower-intermediate participants concluded that the most involving task (sentence writing) enhanced more vocabulary gain and retention than the more involving task (gap filling) which in turn led to more vocabulary gain and retention than the least involving task (reading or video comprehension).

4.2. Vocabulary gain from different types of input at lower-intermediate level

The second research question of the study investigated whether vocabulary tasks with different input modalities (audiovisual vs. written) yield differences in vocabulary gain at lower-intermediate level in terms of productive and receptive word knowledge. When more involving tasks (sentence writing and gap filling) were compared with their equivalents, it was concluded that the pairing of the lower involvement task (gap filling) with the written input and the pairing of the higher involvement task (sentence writing) with the audiovisual input are both effective in promoting the retention of form and meaning.

The learners assigned to the SW task watched a video podcast and became familiar with the target vocabulary through visual cues. As Marefat and Hassanzadeh (2014) suggested, audiovisual materials offer abundant verbal and pictorial cues that can help L2 learners deduce the meaning of unfamiliar words. Mármol and Sánchez-Lafuente (2013) also justified the efficacy of sentence writing in increasing productive word knowledge by proposing that these tasks by nature require employing productive skills. Thus, the type of knowledge they reinforce inherently can offer an explanation for their success on productive vocabulary post-tests.

As for receptive word knowledge, it was concluded that both sentence writing tasks are equally effective in fostering lexical acquisition. Therefore, it can be implied that input modality does not interfere with receptive word knowledge when sentence writing tasks are used at a lower-proficiency level. What is more influential is the deep semantic processing that these tasks induce due to their inherent nature. As Craik and Lockhart (1972) suggested, if L2 learners elaborate on new words more semantically at a deeper level, they tend to recall their meanings more easily. Therefore, the nature of sentence writing tasks largely accounts for the successful retrieval of the target words among the lower-intermediate learners rather than the effect of input type.

With respect to the effects of the written input, it positively affects incidental vocabulary learning when it is consolidated with a gap filling task at a lower language level. As Vidal (2011) claimed, reading is a fundamental source of lexical uptake for less proficient L2 learners because speech segmentation in listening tends to be too complicated for them. Because listening entails real-time processing, learners cannot linger on the words to reflect upon their meanings. However, because the learners in the FR group were provided with a reading passage, they were able to see the target words in their written forms. This enabled them to dwell on meaning and create stronger form-meaning links.

For most L2 learners, it is often problematic to recognize or segment lexical items when they are exposed to spoken discourse (Vandergrift, 2004; van Zeeland, 2013). Besides, when they have trouble understanding the content, they often direct their working memory to comprehending the material rather than focusing on individual words. As a result, their cognitive resources might not be fully available to deal with the unknown words in the discourse (Boers, Eyckmans & Godfroid, 2007). What is more, knowledge of meaning is more difficult to be attained through listening (van Zeeland & Schmitt, 2013) and lexical items acquired by means of listening are more likely to be forgotten by less proficient L2 learners (Vidal, 2011). Considering all these factors, the failure of the FW task in the lower-proficiency group seems to be of no surprise.

4.3. Vocabulary gain from variously loaded tasks at upper-intermediate level

The third research question investigated whether vocabulary tasks with higher levels of involvement load result in more vocabulary gain than tasks with lower levels of involvement load at upper-intermediate level in terms of productive and receptive word knowledge. Thoroughly verifying the predictions of the ILH, the results indicated that sentence writing tasks (index = 3) are more profitable than fill-in tasks (index = 2) which in turn are more effective than reading/watching-only tasks (index = 0) in boosting both productive and receptive word knowledge over time.

One reason for the efficacy of sentence writing tasks at upper-intermediate level might stem from the link between L2 proficiency and cognitive awareness. Research has shown that learners of high L2 proficiency tend to develop and employ more cognitive strategies since linguistic knowledge is the key for such strategies to emerge and operate effectively (Baker, 2005; Griffith & Ruan, 2005). When L2 learners have a better command of the language, they can ease their mental processing and direct their remaining cognitive resources to higher-order skills. That is why the higher-proficiency group gained the most vocabulary through sentence writing although it induced more cognitive load with strong evaluation. Thanks to their grammatical competence and L2 vocabulary size, these learners could overcome the cognitive demand and linguistic difficulty of their task.

On the other hand, there is one notable finding worth pointing out. The FW group obtained an equal amount of productive word knowledge as the SW group and an equal amount of receptive word knowledge as the SR group on the immediate post-tests. Compared to the lower-intermediate group, the upper-intermediate learners assigned to the FW task made considerably more initial vocabulary gain owing to their listening ability. Such learners with higher proficiency possess a larger bulk of lexical knowledge and thus manage to parse the running speech more easily for word-meaning interpretation (Vidal, 2011). This advantage is often unavailable for learners of lower language proficiency. Still, both sentence writing groups outscored the FW group on the delayed post-tests. Thus, it is clear that the higher involvement load of sentence writing still overshadows the efficacy of gap filling over time.

In line with the ILH, the lowest vocabulary gain was obtained by the RO and WO groups at upper-intermediate level. Still, their tasks resulted in more receptive word knowledge than productive word knowledge. This suggests that without any explicit focus on the target words during task implementation, it seemed rather difficult for these learners to acquire the form. However, meaning was

still acquired to some extent even though these learners were not provided with any noticing effects that helped them divert their attention to the target items. Thus, it can be concluded that indirect exposure to unfamiliar words can somewhat lead to their acquisition to a certain degree if learners are willing to employ any inference or guessing strategies.

For the higher-proficiency group, the results entirely validated the hypothesis by stressing the superiority of sentence writing tasks over gap filling tasks followed by reading/watching-only tasks in increasing both productive and receptive word knowledge. In general, the findings emphasized the strength of the evaluation component in task layout.

4.4. Vocabulary gain from different types of input at upper-intermediate level

The final research question explored whether vocabulary tasks with different input modalities (audiovisual vs. written) yield differences in vocabulary gain at upper-intermediate level in terms of productive and receptive word knowledge. The results showed that the audiovisual input resulted in more productive word knowledge across the sentence writing tasks, whereas the written input led to slightly more productive vocabulary retention across the fill-in tasks. When the retention of meaning was considered, the audiovisual input was more effective than the written input with the gap filling task, while both types of input were equal with the sentence writing task in terms of their contribution to receptive word retention.

The findings of the upper-intermediate learners suggested that the audiovisual input seems to yield more vocabulary gain than the written input. Several explanations can be made to account for its superiority. Vandergrift (2004) asserted that the video input provides L2 learners with a great number of pictorial aids that activate their schemata and build up their background knowledge (i.e., top-down processing). Through images, learners can contextualize the information they are attending to. This can reduce the cognitive load on their attentional mechanisms and divert their mental resources to other features of spoken discourse such as segmenting the speech (i.e., bottom-up processing). Thus, they can comprehend the spoken text better and infer the meaning of unknown vocabulary in the input. However, these visual clues are not present in the written input. That is why most of the upper-intermediate learners who read a passage acquired fewer target words than those who watched a video podcast.

A notable distinction emerged between the two proficiency levels when gap filling tasks were compared based on their input presentation. While the FW task was effective among the higher-proficiency learners, it was unsuccessful within the lower-intermediate group. This might stem from the way input modality interacts with L2 proficiency. The FW task required the learners to listen carefully in order to identify the target words. As spoken discourse involves continuous speech (i.e., no clear-cut boundaries), it necessitates instant processing. Thus, listeners basically have less time to comprehend the input. Besides, speech segmentation conveys a particular difficulty for L2 learners since they often miss the upcoming or forget the preceding parts of the listening text (Goh, 2000). When a lower level of proficiency is added to the picture, problems might occur with the acquisition of new words (Vidal, 2011). That might be why the more competent learners in this study attained more vocabulary knowledge than the less proficient ones although both groups performed the same type of task (FW).

Although Hulstijn (2003) asserted that “it is elaboration of or involvement in the lexical information being processed rather than any of the factors per se that determines retention” (p. 364), the present study suggested that input modality is an important variable that influences incidental vocabulary learning. Based on Laufer and Hulstijn’s (2001) claim that tasks constituted with the same level of involvement load are equally effective in enhancing L2 word retention, the tasks in this study were expected to result in similar word gain. However, it was concluded that the gap filling task with the written input and the sentence writing task with the audiovisual input seemed to be more efficient among

the lower-intermediate learners. Similarly, the tasks created with the audiovisual input yielded more vocabulary gain among the higher-proficiency group. All these findings contradict with Laufer and Hulstijn's (2001) prediction on identically loaded tasks leading to similar amounts of lexical uptake. It is evident that task efficacy is influenced by input type which in turn is affected by L2 proficiency. Therefore, input modality might be treated as a fourth component to the ILH when task design is considered.

5. Conclusions

Fully concurring with the predictions of the ILH, the present study favored the tasks with the highest level of involvement (sentence writing) over the tasks that are less involving (gap filling) or induce no involvement (reading alone or video comprehension). Concerning this finding, it can be inferred that strong evaluation has a much stronger impact on the initial gain and retention of new words.

In this study, the learners acquired varying amounts of vocabulary gain from tasks that shared an identical index of involvement. This finding contrasts with Laufer and Hulstijn's (2001) claim for equally loaded tasks that lead to similar vocabulary knowledge. Since this finding applies to both levels, the different word gains can be attributed to the effects of different input types.

The results of the lower-intermediate learners showed that the written input is more effective with a gap filling task, while the audiovisual input facilitates the learning of target words with a sentence writing task. As for the upper-intermediate learners, the findings pointed out the superiority of the audiovisual input. Thus, the present study highlighted the distinctive role that input mode plays in enhancing lexical uptake, especially productive word knowledge, when it interacts with task type and L2 proficiency.

This study offers valuable pedagogical implications for language teachers, material developers and teacher educators. The findings suggest that audiovisual resources can be an effective way of teaching vocabulary. Similarly, engaging L2 learners in vocabulary tasks that instruct them to use the newly encountered words can make them notice these items and thus yield more vocabulary gain. For that, material developers should be encouraged to devise such tasks more. Finally, this study can help prospective language teachers understand the value of word-focused tasks and gain a better insight into task design. Similarly, teacher educators can refer to the findings while explaining their students why they should choose particular tasks but not others for vocabulary teaching.

Acknowledgements

This study is based on the first author's Master's thesis "The Effects of Task Involvement Load and Input Type on Foreign Language Vocabulary Learning" submitted to the Institute of Educational Sciences, Marmara University. The second author is the thesis supervisor.

References

- Anderson, N. J. (1991). Individual differences in strategy use in second language reading and testing. *Modern Language Journal*, 75, 460-472.
- Baker, L. (2005). Developmental differences in metacognitive knowledge: Implications for metacognitively oriented reading instruction. In E. Israel, C. Block, L. Bauserman, & K. Kinnucan-

- Welsch (Eds.), *Metacognition in literacy learning: Theory, assessment, instruction, and professional development* (pp. 61-79). London: Routledge.
- Barcroft, J. (2002). Semantic and structural elaboration in L2 lexical acquisition. *Language Learning*, 52, 323-363.
- Boers, F., Eyckmans, J., & Godfroid, A. (2007). Catering for limited processing capacity to foster incidental vocabulary uptake. In K. Pelsmaekers & C. Rollo (Eds.), *Economically speaking: Essays in honour of Chris Braecke* (pp. 169-187). Antwerp: Garant Publishers.
- Brown, R., Waring, R., & Donkaewbua, S. (2008). Incidental vocabulary acquisition from reading, reading-while-listening, and listening to stories. *Reading in a Foreign Language*, 20, 136-163.
- Craik, F., & Lockhart, R. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behaviour*, 11, 671-684.
- Craik, F., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, 104, 268-294.
- Fatalaki, J. A. (2014). Involvement load hypothesis: Word meaning retention across oral and written task types. *International Letters of Social and Humanistic Sciences Online*, 37, 29-45.
- Folse, K. S. (2006). The effect of type of written exercise on L2 vocabulary retention. *TESOL Quarterly*, 40(2), 273-293.
- Goh, C. (2000). A cognitive perspective on language learners' listening comprehension problems. *System*, 28, 55-75.
- Green, J. M., & Oxford, R. L. (1995). A closer look at learning strategies, L2 proficiency, and gender. *TESOL Quarterly*, 29, 261-297.
- Griffith, P. L., & Ruan, J. (2005). What is metacognitive knowledge and what should be its role in literacy instruction. In E. Israel, C. Block, L. Bauserman, & K. Kinnucan-Welsch (Eds.), *Metacognition in literacy learning: Theory, assessment, instruction, and professional development* (pp. 3-18). London: Routledge.
- Hassanzadeh, M. (2016). Experimenting the influence of input modality on involvement load hypothesis. *International Journal of Social, Behavioural, Educational, Economic, Business and Industrial Engineering*, 10(6), 2043-2048.
- Horst, M., Cobb, T., & Meara, P. (1998). Beyond A Clockwork Orange: Acquiring second language vocabulary through reading. *Reading in a Foreign Language*, 11(2), 207-223.
- Huckin, T., & Coady, J. (1999). Incidental vocabulary acquisition in a second language. *Studies in Second Language Acquisition*, 21(2), 181-193.
- Hulstijn, J. H. (2003). Incidental and intentional learning. In C. J. Doughty & M. H. Long (Eds.), *Handbook of second language acquisition* (pp. 349-381). Oxford: Blackwell.
- Hulstijn, J., & Laufer, B. (2001). Some empirical evidence for the involvement load hypothesis in vocabulary acquisition. *Language Learning*, 51(3), 539-558.
- Jing, L., & Jianbin, H. (2009). An empirical study of the involvement load hypothesis in incidental vocabulary acquisition in EFL listening. *Polyglossia*, 16, 1-11.
- Keating, G. (2008). Task effectiveness and word learning in a second language: The involvement load hypothesis on trial. *Language Teaching Research*, 12(3), 365-386.

- Kim, Y. (2011). The role of task-induced involvement and learner proficiency in L2 vocabulary acquisition. *Language Learning*, 61(1), 100-140.
- Krashen, S. D. (1993). The case for free voluntary reading. *The Canadian Modern Language Review*, 50(1), 72-82.
- 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.
- Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics*, 22, 1-26.
- Lin, L. F. (2010). English learners' incidental vocabulary acquisition in the video-based CALL program. *The Asian EFL Journal*, 12(4), 51-66.
- Maleki, N. A. (2012). The effect of the involvement load hypothesis on improving Iranian EFL learners' incidental vocabulary acquisition in listening comprehension classes. *Australian Journal of Basic and Applied Sciences*, 6(9), 119-128.
- Marefat, F., & Hassanzadeh, M. (2014). Vodcast: A breakthrough in developing incidental vocabulary learning. *Iranian Journal of Applied Linguistics*, 17(2), 27-58.
- Mármol, G. A., & Sánchez-Lafuente, A. A. (2013). The involvement load hypothesis: Its effect on vocabulary learning in primary education. *RESLA*, 26, 11-24.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review*, 63(1), 59-81.
- Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). New York, NY: Cambridge University Press.
- Paribakht, T. S., & Wesche, M. (1993). Reading comprehension and second language development in a comprehension-based ESL program. *TESL Canada Journal*, 11(1), 9-29.
- Paribakht, T. S., & Wesche, M. (1997). Vocabulary enhancement activities and reading for meaning in second language vocabulary development. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition: A rationale for pedagogy* (pp.174-200). New York: Cambridge University Press.
- Pigada, M., & Schmitt, N. (2006). Vocabulary acquisition from extensive reading: A case study. *Reading in a Foreign Language*, 18(1), 1-28.
- Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 129-158.
- Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge: Cambridge University Press.
- Swain, M. (1985) Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 235-256). New York: Newbury House.
- Vandergrift, L. (2004). Listening to learn or learning to listen? *Annual Review of Applied Linguistics*, 24, 3-25.

- van Zeeland, H. (2013). Vocabulary and listening. In C. A. Chapelle (Ed.), *The Encyclopaedia of Applied Linguistics* (pp. 1-6). Australia: Wiley-Blackwell.
- van Zeeland, H., & Schmitt, N. (2013). Incidental vocabulary acquisition through L2 listening. *System*, 41, 609-624.
- Vidal, K. (2011). A comparison of the effects of reading and listening on incidental vocabulary acquisition. *Language Learning*, 61(1), 219-258.
- Waring, R., & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language*, 15(2), 130-163.
- Webb, S., & Rodgers, M. P. H. (2009). The lexical coverage of movies. *Applied Linguistics*, 30(4), 407-427.

Görev kaynaklı katılım yükü ve veri biçiminin rastlantısal kelime öğrenimi üzerindeki etkisi

Öz

Yeterince kelime bilmek ikinci dil öğrenenlerin dili anlaması ve kullanabilmesi için çok önemlidir. Bu nedenle, araştırmaların büyük bir kısmı yeni kelimeler edinebilmek için en etkili yolları bulmayı amaçlamıştır. Görev Kaynaklı Katılım Yükü Hipotezi (Laufer & Hulstijn, 2001) öğrenenin daha etkin katılımı sağlanırsa kelime aktivitelerinin verimli olacağını belirtir. Bu çalışma aynı görev kaynaklı katılım yükü fakat farklı veri biçimlerine (yazılı ya da görsel-işitsel) sahip aktivitelerin farklı dil seviyelerinde kelime öğrenimini nasıl etkileyeceğini araştırmıştır. İngilizce'yi yabancı dil olarak öğrenen 236 (alt-orta ya da üst-orta seviyedeki) Türk öğrenci üç farklı görev kaynaklı katılım yükü ve iki farklı veri biçimine sahip altı kelime aktivitesini gerçekleştirdiler. Aktivitelerden dördü bir metni okuduktan ya da bir videoyu izledikten sonra sekiz hedef kelime ile boşluk doldurma ya da cümle yazma içerirken, diğer ikisi sadece metni okumayı ya da videoyu anlamayı içerdi. Hedef kelimelerin yazılışını ya da anlamını vermeyi gerektiren kelime son testleri ile üretimsel ve algısal kelime bilgisi test edildi. Kelimelerin doğru yazılış yüzdesi ve verilen doğru tanımlar sayılarak veriler elde edildi. Sonuçlar algısal kelime bilgisi için cümle yazmanın (daha yüksek görev katılım yüklü) boşluk doldurmadan (daha düşük görev katılım yüklü) veri biçimi fark etmezsiniz her iki dil seviyesi için de daha etkili olduğunu gösterdi ama görsel-işitsel veri biçimi daha üst seviyedeki öğrenciler arasında algısal kelime bilgisine daha fazla katkıda bulundu. Üretimsel kelime bilgisine gelince boşluk doldurmanın yazılı veri ve cümle yazmanın da görsel-işitsel veri ile birleşimi her iki seviyede de daha etkili oldu. Bu bulgular Görev Kaynaklı Katılım Yükü Hipotezi'nin varsayımlarını algısal kelime bilgisi için kısmen doğrularken, sonuçlar veri biçiminin üretimsel kelime bilgisindeki rolüne de dikkat çekmiştir.

Anahtar sözcükler: Rastlantısal kelime öğrenimi; görev kaynaklı katılım yükü; aktivite verimliliği; veri biçimi; üretimsel ve algısal kelime bilgisi

AUTHOR BIODATA

Cansu Kıvrak got her Master of Arts degree from the Department of Foreign Language Education at Marmara University and is currently working as a lecturer at Yalova University, School of Foreign Languages.

Dilek Uygun Gökmen holds a Ph.D. in linguistics. She is an assistant professor at Marmara University, Faculty of Education, Department of English Language Teaching. Her research areas cover both theoretical and applied linguistics with particular focus on the domains of syntax, morphology and second language acquisition of vocabulary.