THE EFFECTS OF TASK INDUCED INVOLVEMENT LOAD HYPOTHESIS ON TURKISH EFL LEARNERS' INCIDENTAL VOCABULARY LEARNING [1]

(Research article)

(Corresponding author)
Mehtap Yorgancı https://orcid.org/0000-0002-2018-3153
Coordinatorship of Foreign Languages, Konya Technical University, Konya, Turkey
myorganci@ktun.edu.tr

Gonca Subasi https://orcid.org/0000-0001-7049-5940
ELT Department, Anadolu University, Eskisehir, Turkey
goncas@anadolu.edu.tr

Biodata:

Mehtap Yorgancı is an English language instructor at Konya Technical University, Konya, Turkey. She carries out studies in language learning strategies, implicit and explicit vocabulary teaching, teaching pronunciation, and using technology in language classrooms.

Gonca Subaşı is an assistant professor doctor in the ELT Department at Anadolu University, Turkey. Her research interests are teaching writing skills, vocabulary acquisition, affective factors in language teaching, language testing and evaluation, and language teacher education.

[1]This paper is based on the master's thesis titled 'The Effects of Task Induced Involvement Load Hypothesis on Turkish EFL Learners' Incidental Vocabulary Learning' of the first author.

THE EFFECTS OF TASK INDUCED INVOLVEMENT LOAD HYPOTHESIS ON TURKISH EFL LEARNERS' INCIDENTAL VOCABULARY LEARNING

Mehtap Yorgancı

myorganci@ktun.edu.tr

Gonca Subasi

goncas@anadolu.edu.tr

Abstract

Recently vocabulary studies have mainly focused on two forms of vocabulary acquisition: incidental and intentional vocabulary acquisition. For incidental vocabulary acquisition, Task-induced Involvement Load Hypothesis (TILH) was put forward by Hulstijn and Laufer (2001) to investigate the vocabulary tasks by comparing their levels of involvement load to each other. To test this hypothesis, the current study utilized six different vocabulary tasks with varying levels of involvement load. On the other hand, in order to investigate the task type effect, each task with another task from the other task type group was compared. The last part of the study was designed specially to test the task type effect which was neglected by the hypothesis as the hypothesis suggested that only involvement load levels affect the results. The findings concluded that different involvement load levels yielded varying results most of which provided support for the hypothesis. However, task types did not provide evidence in favour of the hypothesis by not leading to similar results for the tasks who shared the same involvement load index. The study concluded with some pedagogical implications and suggestions for further studies.

Keywords: incidental vocabulary, TILH, Task-induced Involvement Load Hypothesis (TILH), Turkish EFL prep students, vocabulary task type effect.

1. Introduction

Vocabulary is a need for all language learners and one of the biggest challenges they face in academic settings. Wilkins (1972) drew attention to the significance of vocabulary by stating that a student may convey little without grammar but nothing without vocabulary. As Folse (2006) suggests for a successful performance in all four skills including reading, having a great deal of vocabulary is necessary. Each student needs to achieve a quality vocabulary in the reading curriculum. One of the reasons that students have difficulty in reading is that they do not have a functional vocabulary for reading. Thus, in EFL teaching and academic studies, the major goal should be enriching and developing learners' vocabulary and finding the most efficient ways for it.

All vocabulary teaching techniques and ways have a main role in vocabulary instruction as they are an open door to vocabulary acquisition. Incidental learning, intentional learning, implicit learning, and explicit learning are some of the most known and used vocabulary learning forms. After using incidental learning in vocabulary teaching, some theories and hypotheses were put forward. One of the most important of them was Depth of Processing Theory proposed by Craik and Lockhart in 1972 and it was criticized for the lack of a clear definition of the level of processing. As a result of the deficiencies, Task-induced Involvement

Load Hypothesis (TILH) was proposed by Laufer and Hulstijn (2001) to provide a more observable and measurable definition.

Because the TILH studies in the literature (Hulstijn, Hollander, and Greidanus, 1996; Wesche and Paribakht, 2000; Hulstijn and Laufer, 2001; Rott, Williams, and Cameron, 2002; Folse, 2006; Sonbul and Schmitt, 2009; Rassaei, 2015; Karalık, 2016; Hazrat, 2020; Teng and Zhang, 2021; Ehsani and Karami, 2022; and Çekiç, 2022) and the hypothesis itself put forward some valuable benefits, the current study was designed within the framework of TILH in an attempt to shed light on the effects of TILH on the EFL prep students' incidental vocabulary learning.

2. Review of Literature
2.1. Task Induced Involvement Load Hypothesis

TILH is used to find the involvement load of vocabulary tasks. As TILH suggests that the higher load of involvement lead to higher vocabulary gain (VG) and vocabulary retention (VR), these involvement loads might be taken into consideration while designing vocabulary tasks. To sum up, teaching vocabulary incidentally using TILH framework might offer some benefits to the language teachers including saving time by combining different ways such as teaching vocabulary while reading.

TILH has three components: need, search, and evaluation. All components have three degrees: absent, moderate, and strong (absent is marked as 0 or -; moderate is marked as 1 or +; and strong is marked as 2 or ++). The moderate and strong degrees for need component change according to type of motivation of the students: extrinsic motivation or intrinsic motivation. As cognitive dimensions, search and evaluation are contingent upon the form meaning relationship (Hulstijn and Laufer, 2001). Search component is marked according to whether the meaning is provided to students or students find it themselves. Evaluation component is more related to an assessment for the appropriate meaning among other meanings of that word and the context of the words.

In incidental vocabulary learning, TILH holds a crucially important place. As a consequence of limited classroom time, the advantages of incidental vocabulary learning were pointed out. And, TILH provides the chance of teaching vocabulary through receptive skills. The vocabulary learning occurs naturally and incidentally for learners. This study aimed at turning these benefits into advantage in the language classrooms. So that, the reading and vocabulary activities may be organised or designed accordingly.

TILH is constructed on some assumptions: (1) presence or absence of the components of need, search, and evaluation determine the level of the retention of the incidentally acquired words; (2) words with higher TILL (Task-induced Involvement Load Level) are retained better than the words with lower TILL. Thus, studies came up with the conclusion that TILH should be taken into account while preparing incidental vocabulary tasks in a specific context. For all these reasons, the current study is conducted to test the effects of TILH on Turkish EFL learners' incidental vocabulary learning.

2.2. Empirical Studies on TILH

As the first research study on TILH, Hulstijn and Laufer (2001) tested their hypothesis with a study in which they compared three tasks. The study is conducted with 186 students in two countries. The aim of the study was to test their own hypothesis, which has the assumption of that the tasks with higher involvement loads lead to higher VG and VR. In this study, the tasks had different involvement load indexes and had different levels for each component (need, search, and evaluation). The effect of involvement load on the retention of ten English words was investigated. The results of this study showed that the tasks with higher involvement load

led to better incidental vocabulary learning. Therefore, the results of this study were compatible with what the hypothesis put forward. Beal (2007) conducted a study using tasks with varying TILLs to test TILH. A short story reading text was used with some unfamiliar words selected for the students and used under four conditions: low, glossary provided; moderate, multiple choice glossary; high, dictionary-based sentence writing; and control, reading only. The findings supported what TILH suggests.

Keating (2008) also tested whether VG and VR were contingent upon tasks' involvement load index as claimed by Laufer and Hulstijn (2001) or not. Seventy-nine beginner level students participated in the study to complete three tasks which have varying levels of involvement load (mental effort). These are: only reading comprehension (no mental effort), providing TWs along with reading comprehension (moderate effort), and formulating original sentences (strong effort). In parallel with what TILH suggests, the task with the highest involvement load led to higher retention than the others, and the moderate level task led to higher retention than the lowest level task. Kim (2008) also conducted a similar study to compare two tasks with the same TILL. However, that study only compared them to see whether the tasks led to similar amount of VG or not. The results showed that two tasks with the same TILLs may lead to a similar amount of incidental vocabulary learning.

On the other side, Zou (2017) conducted a study to compare tasks with the same TILLs to claim that evaluation component should be given another load degree. Although these two tasks (sentence writing and composition writing) seemed that they had the same level, evaluation component should be reconsidered and be given another degree: very strong evaluation. Hazrat (2020) also came up with the conclusion that evaluation factor needs to be given four degrees rather than three concerning its effectiveness for vocabulary learning in her study with ten groups of intermediate level learners. Hazrat (2020) continued by stating that the search component should not have a predetermined degree of prominence and needs to be evaluated based on its relationship with the type of evaluation component with which it is combined in the vocabulary task. Teng and Zhang (2021) investigated the effects of four different tasks which had different TILLs (reading; reading + gap-fill; reading + writing; and reading + writing with the use of a digital dictionary) and supported TILH with the conclusion that the tasks with higher TILLs led to higher VG and VR compared to others.

Çekiç (2022) compared three conditions (traditional gloss, multiple-choice gloss and no gloss) to find out gloss effects on incidental vocabulary learning. Although the two conditions' results outscored no gloss group. Any significant difference was not found between traditional gloss and multiple-choice gloss groups. However, the multiple-choice group was supposed to get higher scores compared to traditional gloss group as they were required to make an appropriate choice. The multiple-choice group had evaluation component which was not present in tradition gloss condition. Çekiç (2022) concluded that the findings of the study seem to contradict TILH based on the results.

The research studies in the literature related to incidental vocabulary learning and TILH mainly included the studies which utilised a variety of vocabulary tasks which had different involvement load levels. The main aim of these studies was to test only TILH without adding any new dimension to this area. However, TILH is limited to only one factor to determine the effectiveness of vocabulary tasks on vocabulary acquisition and more studies need to be conducted to unearth any possible factors. In the literature, there are some TILH studies (Yaqubi, Rayati, and Allemzade Gorgi, 2012; Sarani, Mousapour Negari and Ghaviniat, 2013; Pourakbari and Biria, 2015; Jones-Mensah, Tabiri, Fenyi, Kongo and Amexo) which took task type effect into regard. However, these studies were all conducted in other countries.

In the Turkish context, empirical studies on TILH are limited. Sarbazi (2014) conducted a study in Iran with 30 Turkish EFL learners. He designed three tasks each of which had different TILLs. As the other purpose of the study was to compare the results across gender, the students were assigned to the tasks with the same number of students from each gender. Two-way ANOVA was used for statistical analysis and the results were consistent with what the hypothesis suggested. Any interaction between TILH and gender was not found in the study. In another study, Karalık (2016) compared 139 Turkish ELT students from eight intact groups with four tasks (fill-in by searching TILL: three, fill-in with glossary TILL: two, retelling by searching TILL: four, and retelling with glossary TILL: three) in a Turkish state university. The researcher tried to find if the tasks with higher TILLs yielded higher VG and VR. Another aim of the study was to test if the tasks with the same TILL but having different contributions of the components led to the same results. The results suggested that the tasks with higher involvement loads yielded better results in post-tests. On the delayed post-test, the only significant differences were found between retelling by searching and fill-in groups. The results provided partial support for TILH.

Zou (2017) conducted a study to check the effects of TILH on students' incidental vocabulary learning. For this purpose, 147 participants were assigned to three groups (cloze exercises, sentence writing, and composition writing) randomly. The involvement load of the first task was lower than other two tasks which had the same involvement load. The results showed that the lower involvement load yielded less vocabulary learning. However, to the contrary to TILH there was a statistically significant difference between sentence writing and composition writing group. TILH put forward that two tasks with the same involvement load are expected to result in similar VG and VR. Based on these results, Zou (2017) claimed that the tasks like composition writing which needed a deeper processing and more involvement should be given another degree for the evaluation component. For example, composition writing might have "very strong" evaluation instead of "strong" evaluation as proposed by TILH.

As seen in the studies above, TILH studies in Turkish context investigated the effects of TILH on VG and VR by only comparing the vocabulary tasks with varying levels of involvement load. However, some studies in the literature (Yaqubi et. al., 2012; Sarani et. al., 2013; and Pourakbari and Biria, 2015) added a new dimension to TILH studies by using input-output or receptive-productive vocabulary tasks and testing the effects of task type which was neglected in TILH studies in Turkish context.

Laufer and Hulstijn (2001) acknowledged that task's efficacy is only determined using TILLs. In other saying, two receptive and two productive tasks yield equal vocabulary learning as long as they share the same TILL. However, as a suggestion for further research, they also claimed that some studies may be designed to see if there is any difference between receptive and productive tasks which are equal in involvement loads. Yaqubi et. al. (2012) came up with the conclusion that the task type whether it was an input or output task had a crucial effect on incidental vocabulary learning. The input tasks were compared with each other and found that the task with higher involvement load led to higher VG and VR. The hypothesis claimed that the tasks with the same involvement load index yield similar results in VG and VR. Sarani et.al. (2013) conducted a study to see task type effect on TILH through reading. For this aim, three receptive and three productive tasks were designed for six groups. Two pairs with the same involvement load (e.g. tasks with involvement load of 1: true-false, receptive task and short response, productive task) gave a contrary result to TILH. Pourakbari and Biria (2015) designed a study with three receptive and three productive tasks with different TILLs. The last research question was asked to see if or not task type would make any difference in incidental

VG and VR. Task type influenced TILH as productive tasks were found to be more beneficial for incidental VG and VR.

In conclusion, previous research findings have shown that TILH has a crucial place in incidental vocabulary teaching. However, TILH is limited to only involvement load levels and does not take any other factor into consideration. Any particular task type – input output or receptive productive – does not make any difference and is not more effective than the other according to TILH as the only factor that affects efficacy of tasks in TILH is how much involvement they require. As a result of this, it is a requirement to conduct more studies focusing on tasks with similar levels of involvement but from different task types. To follow Laufer and Hulstijn's (2001) suggestion to conduct a study in which tasks from different types but having identical involvement loads are examined, this study was designed with three receptive and three productive tasks with different involvement loads. Receptive and productive tasks had involvement loads of 1, 2, and 3 respectively in their task type. On the contrary, each task has a conjugate task with the same load in the other task type. To shed light on this lack in the literature, this study posed the following research questions:

1) On the basis of English receptive vocabulary tasks, will EFL prep learners obtain better gain of lexical items in higher task load conditions compared to lower ones? If so, will the benefits of tasks retain over time?

2) On the basis of English productive vocabulary tasks, will EFL prep learners obtain better gain of lexical items in higher task load conditions compared to lower ones? If so, will the benefits of tasks retain over time?

3) On the basis of English receptive and productive vocabulary tasks with the same levels of involvement index, will EFL prep learners obtain the same gain and retention of the lexical items on both types of tasks?

3. Methodology

The aim of the study was to investigate the effects of task-induced involvement load on incidental vocabulary acquisition of EFL learners through different vocabulary tasks which were designed taking TILH framework into consideration. To this end, different groups of students were assigned to different incidental vocabulary tasks with different involvement loads.

For the design of the current study, non-control grouped quasi-experimental research design was chosen as the study lacked a pre-test but comprised of two post-tests (immediate and delayed post-tests). Moreover, the present study was designed without a control group but six different experimental groups to test the effect of various tasks on students' incidental vocabulary learning. As Creswell (2005) states, researchers generally use intact groups either because of the availability of the participants or because of the setting which does not let creating artificial groups. Similarly, in this study forming groups of students was not applicable. Instead, the classes were taken as intact groups; therefore, a random assignment of the subjects was unlikely.

3.1. The Setting of the Study and the Participants

The study was conducted at the School of Foreign Languages of a foundation university in one of the cities of Turkey. The students who do not pass the English proficiency exam conducted by the universities at the beginning of the academic year are obliged to have a one year Intensive English Programme at universities' School of Foreign Languages in Turkey. The participants of the study were 122 Turkish EFL students who were having A2 level intensive English course during the study. All these intact classes were experimental groups.

Convenience sampling was preferred for the selection of the participants. Hence, all A2 level students were asked to participate in the current study.

Before implementations, the ethic committee of the university was applied for the informed consent. After that, a consent form was collected from all instructors and the students who were willing to participate in the study.

3.2. Instruments

In this study, four instruments were utilised for research purposes. The students were required to read one text and complete its reading comprehension activities which were taken directly from a reading skill book. After the text and reading activities, each group was asked to complete one vocabulary task. The vocabulary tasks were designed differently for each group to measure its effect on students VG and VR scores within TILH framework. Nine target words (TW) in the text were chosen for the study. The TWs should be unknown to the participants. Therefore, TWs were checked by four colleagues. The TWs chosen for the study were forehead, holy, mud, to please, prosperity, to receive, stray, trail, and to worship. The vocabulary activities were designed by the researcher and all activities were checked by four other instructors at the same school and two professors at English Language Teaching department of a state university for their appropriateness for the participants and to check validity. In order to have a robust design, first, expert opinion was gathered for the validity of the instruments utilized in the current study. Then, the reading text's readability was analysed through Flesch Kincaid Grade Level to achieve reliability.

The students were asked to complete a vocabulary task which was assigned to each group randomly for the purpose of the current study. The tasks were true/false, matching with definitions, multiple choice, short response, fill in the blanks, and sentence writing. The tasks were categorized into productive and receptive tasks. Table 1. shows the tasks and their total involvement load indexes.

Table 1. *Total task induced involvement load levels of vocabulary tasks*

	Tasks	Need	Search	Evaluation	Total TILL
Receptive Tasks	True/False (a)	1	0	0	1
	Matching (b)	1	0	1	2
	Multiple Choice (c)	1	1	1	3
Productive Tasks	Short Response (d)	1	0	0	1
	Fill in the Blanks (e)	1	0	1	2
	Sentence Writing (f)	1	0	2	3

As seen in Table 1., the participants were divided into six groups and the groups were categorized according to their task type as receptive and productive. The receptive and productive task type groups both had three different vocabulary tasks. The TILLs of the tasks were designed as to see if there was any difference of the same TIL but different task type had any effect on incidental vocabulary learning. On the other hand, the tasks with the same

involvement but from different task type were designed with the intention of comparing them to each other to see any possible task type effect on incidental learning which was not mentioned in the TILH.

The questions in the vocabulary tasks included the TWs, so the students needed to know the meanings of the TWs in order to give the right answers. The design would help the researcher compare each task with other tasks in its own task type and compare them to their conjugate tasks from the other task type group which had the same TILL. The tasks sharing the same TILL also shared the same allocated time. While completing the task, the students were encouraged to use the glossary provided at the end of the text.

A modified version of VKS (Vocabulary Knowledge Scale) used in Hulstijn and Laufer (2001) was preferred in both immediate and delayed post-tests to measure the VG and VR. The self-reported VKS consists of four items and is shown in Figure 1. below.

	Items	Score
Target word: _____	I can't recall having seen this word before.	0
	I have seen this word before, but I can't remember what it means.	1
	I have seen this word before, and I think it means: _____	2
	I can use this word in a sentence: _____	3

Figure 1. The self-reported modified VKS

To score this modified VKS, the participants did not receive any point when they marked that they did not remember the word; one point was awarded when only the form of the TW was recalled; the students received two points when they provided the Turkish equivalents or English definitions of the TWs; and the students who generated a sentence using the TWs received three points.

The immediate post-test was administered immediately after the students finished reading the passage and completed reading comprehension question and their vocabulary tasks. The delayed post-test was administered three weeks later. The delayed post-test was the same with the immediate post-test. The only difference was the order of appearance of the TWs to prevent the students' from remembering them in that order and giving their answers accordingly.

3.3. Analysis of the Data

Two vocabulary tests were used to compare the results of the effects of two task types. The students' scores on immediate post-test was compared to each other to measure their immediate VG. On the immediate post-test, the students were asked to provide Turkish equivalents, English synonyms/definitions of the TWs or generate a meaningful sentence using the TWs. There were four options for each TW and the students were to put a tick next to only one of them. The scores of the options were 0, 1, 2, and 3 respectively. In a similar way, the same procedure was applied for the delayed post-test which was unannouncedly administered three weeks later. The scores of the students from delayed post-test were compared to test VR. The data was analysed using SPSS 22.

For the research purposes of the current study, three research questions were posed with different purposes and a different design from many of the TILH studies in the literature. A summary of the design of the current study is presented in Table 2.

Table 2. *A Summary of the research questions and the design*

Research Questions	Method of Analysis	Purpose of the Research Question
1. On the basis of English receptive vocabulary tasks, will EFL prep learners obtain better gain of lexical items in higher task load conditions compared to lower ones? If so, will the benefits of tasks hold up over time?	one-way ANOVA and paired t-test	to test the TILH among the receptive vocabulary tasks and to find out the effect of time interval
2. On the basis of English productive vocabulary tasks, will EFL prep learners obtain better gain of lexical items in higher task load conditions compared to lower ones? If so, will the benefits of tasks hold up over time?	one-way ANOVA and paired t-test	to test the TILH among the productive vocabulary tasks and to find out the effect of time interval
3. On the basis of English receptive and productive vocabulary tasks with the same levels of involvement index, will EFL prep learners obtain the same gain and retention of the lexical items on both types of tasks?	independent samples t-test	to test any possible task type effect between the vocabulary tasks with the same TILL which was neglected in the TILH

4. Results and Discussion

To serve the purpose of the current study, three research questions were posed. The research questions were addressed to find out the vocabulary gain and vocabulary retention of Turkish EFL learners through incidental vocabulary learning within the construct of TILH and vocabulary task types.

Six different vocabulary tasks from two different task types were utilised to see the effect of TILL on EFL learners' incidental vocabulary acquisition. A reading text was chosen to operationalise different indexes of involvement loads. Before analysing the data, the distribution of the scores of six groups from both immediate and delayed post-test were examined. The results of the normality tests, skewness and kurtosis values showed that the scores were normally distributed. Therefore, parametric analyses (one-way ANOVA, independent samples t-tests, and paired t-tests) were utilized for the data analysis of the current study. For the first two research questions, a comparison was made in order to find out the most effective vocabulary tasks in each task type. To answer the last research question, three different comparisons were made, and each task was compared to its conjugate task which shared the same TILL from the other task type group.

4.1. Tasks with Different Involvement Load Levels

The comparisons will be made between tasks with different involvement loads. However, each task will be compared to the other tasks in their own task type on the immediate and delayed post-tests.

The scores of all receptive and productive task groups are presented in Table 3. Table 3. is used to compare the mean scores of all groups, and also to compare the highest and the lowest scores of the groups.

Table 3. *A summary of mean scores of all groups*

	Immediate	Delayed
R1 (True/False)	12.60	8.85
R2 (Matching with Definitions)	18.05	12.55
R3 (Multiple Choice)	9.42	7.63
P1 (Short response)	15.35	10.04
P2 (Fill-in)	14.89	11
P3 (Sentence Writing)	15.33	11.43

4.1.1. Receptive Vocabulary Tasks

The results on both immediate and delayed post-tests provided support for TILH partially. The higher TILLs yielded better VG and VR in most of the statistical analyses.

Table 4. *Immediate vocabulary gain scores of receptive tasks group*

	N	M	SD	Min.	Max.
R1 (True/False)	20	12.60	4.096	5	19
R2 (Matching with Definitions)	20	18.05	4.850	11	26
R3 (Multiple Choice)	19	9.42	3.820	3	16

Table 5. *Vocabulary retention scores of receptive task groups*

	N	M	SD	Min.	Max.
R1 (True/False)	20	8.85	2.834	3	14
R2 (Matching with Definitions)	20	12.55	3.395	7	20
R3 (Multiple Choice)	19	7.63	2.985	3	14

As seen in table 4. and table 5., on both the immediate and delayed post-tests of receptive group, contrary to TILH, R2 with a TILL of 2 received the highest scores which was followed by R1 and R3 groups. To find out the difference between the receptive task groups, one-way ANOVA was conducted for both tests.

Table 6. *One-way ANOVA for immediate post-test scores of receptive task groups*

	Sum of Squares	df	Mean Square	F	Sig.	Sig. Difference
Between Groups	746.025	2	373.013	20.312	.000	R1-R2; R2-R3
Within Groups	1028.382	56	18.364			
Total	1774.407	58				

Table 7. *One-way ANOVA for delayed post-test scores of receptive task groups*

	Sum of Squares	df	Mean Square	F	Sig.	Sig. Difference
Between Groups	258.181	2	129.090	13.590	.000	R1-R2; R2-R3
Within Groups	531.921	56	9.499			
Total	790.102	58				

Table 6. and table 7. show that a significant difference in terms of vocabulary gain and vocabulary retention of receptive tasks was indicated separately in the results of one-way ANOVA (F=20.312, p<.05 for immediate post-test) (F=13.590, p<.05 for delayed post-test). In order to detect which groups differed from each other significantly, post-hoc Tukey tests were conducted for both of the tests. A significant difference was found between R1 (M=12.60, SD=4.096) and R2 (M=18.05, SD=4.850) groups and R2 (M=18.05, SD=4.850) and R3 (M=9.42, SD=3.820) groups based on the results of immediate post-test. And again, for the delayed post-test the results of the post-hoc Tukey test resulted in that a significant difference was found between R1 (M=8.85, SD=2.834) and R2 (M=12.55, SD=3.395) groups and R2 (M=12.55, SD=3.395) and R3 (M=7.63, SD=2.985) groups. Any difference between R1 and R3 groups was not found for both of the tests.

Thus far, the first part of the first research question was tried to be answered. The rest of the first research question was "if so, will the benefits of tasks hold up over time?" To this end, paired samples t-tests were conducted to compare each receptive task. A comparison between immediate and delayed post-tests of each receptive was made and a significant difference was found between all of the tasks' immediate and delayed post-tests.

4.1.2. Productive Vocabulary Tasks

The results of immediate post-test and delayed post-test were compared for VG and VR of the productive task groups. All groups gained the meanings of the target words to some extent. The results were different in productive task groups.

Table 8. *Immediate vocabulary gain scores of productive tasks group*

	N	M	SD	Min.	Max.
P1 (Short response)	23	15.35	4.018	9	22
P2 (Fill-in)	19	14.89	5.363	0	23
P3 (Sentence Writing)	21	15.33	5.228	3	26

As seen in table 8., on the immediate post-test, P1 outscored P3 and P1 group's scores were followed by P3 and P2, respectively. This time, the results on the delayed post-test were different from the results of productive groups' immediate post-test.

Table 9. *One-way ANOVA for immediate post-test scores of productive task groups*

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	2.644	2	1.322	.056	.946
Within Groups	1419.674	60	23.661		
Total	1422.317	62			

Table 9. shows that according to the one-way ANOVA results of the participants, a significant difference between was not found between groups (F=0.56, p>.05). As the results of one-way ANOVA was insignificant, a post-hoc test was not conducted.

Table 10. *Vocabulary retention scores of productive task groups*

	N	M	SD	Min.	Max.
P1 (Short response)	23	10.04	2.585	2	13
P2 (Fill-in)	19	11.00	4.256	3	22
P3 (Sentence Writing)	21	11.43	3.340	3	17

To find out the long-term effect of TILL, a similar statistical analysis was carried out for the delayed post-test data to examine the differences between the groups. The delayed post-test results show in table 10 that all results were in line with TILH. Hence, the highest group was found to be P3 and was followed by P2 and P1. These results showed that, among productive vocabulary tasks, although P1 was found to be the task which provided highest scores on the immediate post-test, time interval affected the situation negatively and on the delayed post-test P1 group was found to be lowest group.

Table 11. *One-way ANOVA for delayed post-test scores of productive task groups*

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	22.218	2	11.109	.958	.390
Within Groups	696.099	60	11.602		
Total	718.317	62			

One-way ANOVA results of the participants did not yield a significant difference between the groups (F=.958, p>.05) in table 11. As the results of one-way ANOVA was insignificant, a post-hoc test was not conducted.

The rest of the first research question was "if so, will the benefits of tasks hold up over time?" For this aim, three paired samples t-tests were conducted to compare the immediate and delayed post-tests of all productive tasks. As a result, a significant difference was found between all post-tests.

4.2. Tasks with the Same Involvement Load Levels

The third purpose of the current study was to investigate whether different tasks with the same TILL from different task types would lead to similar results in VG and VR which made the current study different from other TILH studies in literature. In order to attain this purpose, each task was compared to its conjugate task from the other task types on immediate and delayed vocabulary post-tests. To this end, the statistical analysis was conducted using independent samples t-test.

For this aim, two tasks sharing the same TILL from two task types were compared in terms of VG and VR.

Table 12. *Comparison of immediate vocabulary gain scores of groups with their conjugate tasks*

	M	SD	M	SD	
R1 (True/False)	12.60	4.096	15.35	4.018	P1 (Short response)
R2 (Matching with Definitions)	18.05	4.850	14.89	5.363	P2 (Fill-in)
R3 (Multiple Choice)	9.42	3.820	15.33	5.228	P3 (Sentence Writing)

Table 13. *Comparison of vocabulary retention scores of groups with their conjugate tasks*

	M	SD	M	SD	
R1 (True/False)	8.85	2.834	10.04	2.585	P1 (Short response)
R2 (Matching with Definitions)	12.55	3.395	11.00	4.256	P2 (Fill-in)
R3 (Multiple Choice)	7.63	2.985	11.43	3.340	P3 (Sentence Writing)

As seen in table 12. and table 13., P1 group (M=15.35, SD=4.018) had higher scores than R1 group (M=12.60, SD=4.096); and P3 group (M=15.33, SD=5.228) and R3 group (M=9.42, SD=3.820) had a substantial difference on both the immediate and delayed post-tests. And the results of independent samples t-test indicated a significant difference on immediate post-test

scores. Although P1 group outperformed R1 group and P3 group had higher scores than R3 group on the delayed post-test in terms of VR, the differences were insignificant between the delayed post-test scores. On the other hand, R2 group (M=18.05, SD=4.850) and P2 group (M=14.89, SD=5.363) yielded similar results on the immediate post-test and the independent samples t-test results indicated a significant difference between these two groups. Another similar result was observed between the scores of delayed post-tests of P2 (M=11.00, SD=4.256) and R2 (M=12.55, SD=3.395) groups. Even though the students who completed R2 outperformed the students who completed P2, a significant difference was not found on the delayed post-test scores of two groups in terms of retention.

4.3. Discussion of the Findings

A division of productive and receptive tasks made it possible to compare each task in its own task type as in research question one and research question two. It also facilitated to compare two tasks sharing the same TILL from two different tasks types to find out the task type effect as in research question three. The receptive group tasks required the participants to recognize the form and meaning of the TWs and choose the correct answer by matching, determining if they are true/false, and choosing the meanings in multiple choice questions. However, the productive group tasks required to provide a product by writing a few words to answer questions, fill in the blanks of a text, and generating a meaningful sentence.

The findings of both post-tests implied that involvement load level had an effect on the participants' incidental vocabulary learning to some extent. Not all comparisons yielded the expected results caused by TILH. However, most of the comparisons were in line with TILH.

The reasons of not having the same results in line with the other TILH studies in the literature might include the fact that the students might not take the tasks seriously as they were informed that they were not going to get any score from these tasks. Another reason might be related to difference between the classes. Although all the participants were A2 level during the current study, there were some differences between the classes. The vocabulary tasks were assigned randomly, therefore the results might have been affected from these language level differences. Time limitation was another factor as each quarter at the School of Foreign Languages lasted 8 weeks, the implementation which included the tasks, immediate post-test and three weeks later a delayed post-test was also affected by this restriction.

4.3.1. The effects of tasks with different TILLs

The statistical analysis on both of the post-tests of receptive task groups showed that the groups differed from each other. According to TILH, R3 was expected to get the highest scores. R2 was supposed to yield better results than R1. However, the highest scores belonged to R2, R3, and R1task groups, respectively.

The findings of the statistical analyses conducted for productive task groups indicated that the productive groups differed from each other on both post-tests. As TILH suggested, the highest scores should have belonged to P3 group and the lowest scores belonged to P1 group. However, the results of the immediate post-test showed that P1 group outscored the other groups and the lowest scores were obtained by P2 group. And delayed post-test results provided full support for TILH by having the highest scores from P3, P2, and P1 groups, respectively.

The answers for the first research question provided partial support for TILH contrary to the similar research studies in the literature (Sarani et. al., 2013 and Pourakbari and Biria, 2015). The current study did not provide full support for TILH on both post-tests. Kim (2008) also concluded with partial support to TILH as the task with the highest TILL did better on the post-test. However, the task with the moderate level of involvement load was not found to be superior to the task with the lowest TILL. The task with the highest TILL was found to provide

the lowest scores on both post-tests and this might be because of the fact that adding search component to a task might not provide the expected results.

The second part of the first research question was related to the effect time interval. The results showed that the scores decreased to some extent when the participants' scores of two post-tests were compared. Therefore, it could be stated that three weeks' time interval affected the scores of the participants of the current study negatively as in Behbahani, Pourdana, Maleki, and Javanbakht (2011), Arpaci (2016), and Ehsani and Karami (2022).

On the other hand, the answers for the second research question provided two different results. While the results of the immediate post-test provided partially support for TILH, the results of the delayed post-test provided full support which was only obtained from the comparison scores of the productive task groups' post -tests. Similarly Folse (2006) and Walsh (2009) did not conclude with the results which were in line with TILH. Like Walsh (2009), the current study did not provide any significant difference on the results of one-way ANOVA. The mean scores of the groups were very similar.

The results of delayed post-test score comparisons to find out effects of TILH on word retention provided full support for TILH like many studies in the literature (Hulstijn and Laufer, 2001; Beal, 2007; Keating, 2008; Kim, 2008; Eckerth and Tavakoli, 2012; and Mármol and Sánchez-Lafuente, 2013). All concluded that the tasks with higher involvement loads led to higher VG and VR. In the current study, sentence writing group who has the highest TILL outscored the other two groups, namely short response and fill in the blanks.

As Behbahani et. al., (2011) put forward in their study, it is not a surprising fact to have students who did better on the immediate post-test and then their scores decreased on the delayed post-test. This situation could be associated with negative time interval effect between the two post-tests. Hence, the scores of the participants of the current study negatively might have been affected negatively due to three weeks' time interval.

4.3.2. The effects of tasks with the same TILLs

An attempt was made to find whether the tasks with the same TILLs would yield similar results or not. To this end, the third research question of the current study was posed to find out any possible task type effect on students' VG and VR. The tasks were matched with their conjugate tasks. Each pair was compared to each other on both immediate and delayed post-tests.

The comparison of the first pair (P1 and R1) on the immediate and delayed post-tests showed that productive tasks lead to higher VG and VR on post-tests. The findings are in line with the findings of Ellis and He (1999) who suggested that the students remember productive tasks better than non-productive tasks.

In the current study, as the second pair, P2 and R2 groups were compared. To the contrary of the suggestions of TILH, these two groups did not have similar results on the post-tests. Some studies in the literature provide support for the situation. Laufer (2003) revealed that sentence completion group (TILL:3) had higher scores on the tests than sentence writing group (TILL:3). In Esfahani's study (2012), firstly productive group outperformed the other group in writing test, and then the receptive group did better in the reading comprehension test. As Webb (2005) suggested most of the vocabulary tasks in a classroom setting are receptive tasks. Hence, the students in P2 and R2 groups might be more familiar with the receptive tasks. As a result, the reason of R2 group's having higher scores on both post-tests might be explained. Folse (2006) compared three receptive tasks (cloze exercises) with one productive task (sentence writing). The results showed that receptive task groups outperformed the productive task group.

To the contrary of this fact, some other studies were conducted and provided counterevidence for receptive tasks' superiority.

Laufer and Rozovski-Roitblat (2011) advocated that most of the linguistic resources should be used for productive tasks. Webb (2009) concluded that the students assigned to productive tasks did better on the tests compared to the students assigned to receptive group. Like these studies, in the present study P3 group obtained higher scores than R3 group in the comparison between them as the last pair.

To sum up, the tasks sharing the same involvement load did not lead to similar results in any of the pairs. The findings of the study supported the findings of Yaqubi et. al. (2012) who suggested that other than the involvement index, task type (receptive or productive) has a crucial role in incidental vocabulary learning of EFL learners. Therefore, taking task type effect into consideration while designing vocabulary tasks along with TILH might provide useful insights for scholars and language teachers.

5. Conclusion

For the current study, six different vocabulary tasks with varying total involvement load indexes were designed to conduct the present study which aimed to find out the effects of Task-induced Involvement Load Hypothesis on the incidental vocabulary acquisition of 122 EFL prep students at a private university. A reading text with its nine target words was utilised to test the participants' incidental VG and VR. The text was accompanied first with two different reading comprehension activities and then each group was given a vocabulary task which was specifically designed for that group. To measure VG and VR, unannounced immediate and delayed post-tests were conducted. The scores that participants obtained from these two post-tests were analysed to find out the effects of TILH on the participants' incidental vocabulary acquisition.

In order to answer the first research question which sought whether three receptive tasks with varying levels of involvement load had any effects on students' VG and VR, the scores obtained from immediate and delayed post-tests were compared and it was found that the target words were remembered by most of the participants on both post-tests. Although the results of two post-tests for receptive tasks were similar to each other, these results did not support TILH completely. Increasing the total involvement load indexes did not bring about the expected results as anticipated in the hypothesis which can be seen for multiple-choice group who were supposed to outscore the other two groups. Although the two lowest were not found as expected (R2>R1>R3), the difference between R1 and R3 groups were found to be insignificant on both post-tests. It showed that increasing involvement load levels of all vocabulary tasks might not provide the desired results. Some tasks might be affected by other factors. In order to explore it in detail, more receptive vocabulary tasks with varying TILLs might be compared to each other.

Similar to research question one, the second research question aimed at finding whether three productive vocabulary tasks with different total involvement load indexes had any effects on the participants' VG and VR on the post-tests. The hypothesis put forward that between these three receptive tasks the highest scores should have belonged to P3, the higher scores to P2, and the lowest scores to P1. Contrary to the findings of the immediate post-test, the results of the comparisons supported the hypothesis fully (P3>P2>P1). Although P1 group received the highest scores on immediate post-test, it was the group who obtained the lowest scores on the delayed post-test. This might prove that providing short response to the questions as in P1 group might help students remember the words in their short-term memory. However, it does not help retaining the words in the long term.

The effect of time interval for both receptive and productive vocabulary tasks have also been investigated as a part of the research questions one and two. It was found out that for both groups the three weeks' time interval affected negatively. However, this was an expected result as the students did not receive any treatment or language education related to these TWs.

For the third research question, it was aimed to find out any possible significant difference between the groups who shared the same level of involvement load. To this end, three pairs were compared to each other on VG and VR. All comparisons yielded a significant difference between the groups in each pair on the immediate post-test. However, the differences between the scores of the groups in each pair (P1-R1; P2-R2; P3-R3) on the delayed post-test were found to be insignificant. This might be due to the differences between the levels of being affected by the time interval of both parts of the pairs.

As a result of comparing the tasks to their pairs, the pairs sharing a TILL of 1 and 3, productive vocabulary tasks outperformed the receptive vocabulary tasks. However, the comparison between the tasks sharing a TILL of 2 concluded that receptive task group (R2) did better than productive task group (P2). Although, some studies in the literature like Ellis and He (1999) provided results in support of productive tasks' superiority, the second pair (P2-R2) provided counterevidence in this present study. In fact, the findings might change according to not only the task type but also to other factors because Esfahani (2012) also concluded with firstly the results in favour of productive tasks and then counterevidence to productive tasks. It can be concluded that productive tasks' superiority over receptive tasks might be found in most of the comparisons. However, it would be a good idea to take other factors such as task features and requirements into consideration not to overgeneralize the results. Additionally, Ehsani and Karami (2022) came up with the conclusion that Technique Feature Analysis (TFA) is a more powerful predictor for incidental vocabulary learning than TILH as TILH has many shortcomings and they are compensated for by TFA model.

5.1. Implications

In an attempt to test TILH, six vocabulary tasks with different TILLs were designed. These tasks were categorized into two groups, receptive and productive, both to compare them in their own task type and to compare each task to its conjugate task which shares the same involvement load level in the other task type group. Unlike other TILH studies in the literature, the current study aimed at adding a new dimension to the hypothesis by taking the effects of task type into consideration. Hence, the findings of the study offer some implications for both TILH literature and classroom practices regarding incidental vocabulary acquisition.

Regardless of task type, any vocabulary task should be designed by taking its involvement load index into consideration as in most of the comparisons of the current study, it was found out that the higher TILL both led to higher VG and VR. As many studies in the literature like Yaqubi et. al. (2012), Sarani et. al. (2013), Pourakbari and Biria (2015), and Karalık (2016) suggested, the tasks with higher involvement loads should be selected in order to increase VG and VR.

The present study tested TILH. However, on the other hand, it was found out that making use of vocabulary tasks for incidental learning also helped draw students' attention on the target words. Karalık (2016) and Eysenck (1982) put forward that it was not the willingness of the students but how deeply the word is processed at the first encounter to be able to store the words in the memory successfully. Hence, the vocabulary tasks like the tasks of the current study might be helpful for incidental vocabulary learning. As classroom time is limited to teach everything intentionally, incidental teaching techniques should be preferred.

The reason of not having found similar results in the current study as TILH suggested that the students might be used to doing some specific vocabulary tasks such as matching with definitions and true/false as many course books provide these two tasks mostly in the first levels (A1 and A2). Alavinia and Rahimi (2019) advocated that some other factors related to the students such as attention span, writing skills, and dictionary use might hinder the effect of TILH. In the context of the current study, the students are always encouraged to use a dictionary. However, any training on choosing the best definition for the context is not provided to the students.

The participants of the current study have practice in short response activities mostly and they are mostly asked to answer these questions in the exams of their school. Hence, the attention of the students is generally drawn to short response vocabulary tasks. Writing sentences and paragraphs using the target words studied in the reading passages are postponed until B1 level. Therefore, the students do not get used to writing sentences immediately and it takes more time until they feel comfortable with writing sentences and using the target words in them. As Zou (2017) stated writing exercises help students more in vocabulary learning compared to other vocabulary exercises like cloze exercises as writing exercises require pre-planning and systematic organization which are absent in other vocabulary exercises. It would be a good idea to start writing sentences along with vocabulary teaching in order to have more comfortable students in producing the target language verbally. Zou (2017) added for the reading-based exercises of the teaching materials, writing sentences using the target vocabulary should be attached the necessary importance as the students are supposed to use chunking, pre-task planning, and hierarchical organization for writing. As Ehsani and Karami (2022) suggested, the internal structures of the vocabulary tasks lead to different test results. These structures also identify the TILLs of the vocabulary tasks.

5.2. Limitations of the Present Study

For this purpose of the study, more vocabulary tasks might be designed. The results of the current study may be generalized for the tasks included here. Each task has its own peculiar result on different tests. Hence, for the long-term retention the results of the immediate post-test might be taken into consideration.

The study implemented the study just once in order not to make students be aware of the upcoming tests and the aim of the present study. More implementations of the same design with different reading passages over time might yield different results. However, as the nature of the incidental vocabulary learning having students who knew about the upcoming procedure would not suitable with the nature of incidental teaching.

The vocabulary test scores were not graded by a second professional. Only for the ambiguous answers, expert opinion was gathered. That might have affected some of the results. The findings of the study are limited to this specific context. Different studies with participants from state universities, different departments, backgrounds and with different levels of English might yield different results.

5.3. Suggestions for Further Studies

In the current study, the findings concluded that both receptive and productive tasks might yield more different results than what TILH suggested. Hence, the comparisons of the post-test scores might be taken into consideration as to find out the most useful vocabulary tasks.

Zou (2017) who conducted a study and compared two productive tasks with a TILL of 3 concluded that although sentence writing and composition writing shared the same TILL, composition writing group outperformed the other. Hence, a new degree of evaluation should

be added for new studies. In this study, the productive tasks did not differ from each other much. Therefore, a new design like Zou (2017) might be preferred in the further studies.

Unlike TILH, the current study came up with the conclusion that in all of the comparisons a task type had superiority over the other. TILH suggested that the tasks sharing the equal involvement load levels yield similar results. Therefore, further studies might utilize productive tasks more than receptive tasks.

In the current study, only one delayed post-test was conducted three weeks after the implementation. Another delayed post-test might be conducted more weeks later in order to vocabulary retention in longer time periods.

Many studies in the literature and the current study concluded with some counterevidence to TILH. Although the hypothesis leads to more VG and VR based on TILLs of the vocabulary tasks, the fact that it might not be so effective for all vocabulary tasks should be taken into consideration while designing further studies.

References

Alavinia, P. and Rahimi, H. (2019). Task types effects and task involvement load on vocabulary learning of EFL learners. *International Journal of Instruction, 12*(1), 1501-1516.

Arpaci, D. (2016). The effects of accessing L1 versus L2 definitional glosses on L2 learners' reading comprehension and vocabulary learning. *Eurasian Journal of Applied Linguistics, 2*(1), 15-29.

Beal, V. (2007). *The weight of involvement load in college level reading and vocabulary tasks.* Doctoral dissertation. Canada: Concordia University.

Behbahani, S. M. K., Pourdana, N., Maleki, M., Javanbakht, Z. (2011). EFL task induced involvement and incidental vocabulary learning: Succeeded or surrounded. *International Conference on Languages, Literature and Linguistics. IPEDR Proceedings,* 26, 323-325.

Craik, F. I. and Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*(6), 671-684. Retrieved December 22, 2018, from http://dx.doi.org/10.1016/S0022-5371(72)80001-X

Creswell, J. W. (2005). Educational research: Planning, conducting, and evaluating quantitative and qualitative research (2nd ed.). Upper Saddle River, NJ: Pearson.

Çekiç, A. (2022). Incidental L2 vocabulary learning from audiovisual input: the effects of different types of glosses. *Computer Assisted Language Learning*, 1-28.

Eckerth, J. and Tavakoli, P. (2012). The effects of word exposure frequency and elaboration of word processing on incidental L2 vocabulary acquisition

through reading. *Language Teaching Research*, 16(2), 227-252.

Ehsani, M., & Karami, H. (2022). Comparing the predictive power of involvement load hypothesis and technique feature analysis. *International Journal of Language Studies*, *16*(2).

Ellis, R. and He, X. (1999). The roles of modified input and output in the incidental acquisition of word meanings. *Studies in Second Language Acquisition, 21*(2), 285-301.

Esfahani, F. R. (2012). Impact of vocabulary learning tasks on communicative gains of advanced EFL learners of Persian. *American Journal of Economics,* 14-17.

Eysenck, M.W. (1982). *Incidental learning and orienting tasks. In C. R. Puff (Ed.), Handbook of research methods in human memory and cognition.* New York: Academic Press.

Folse, K. S. (2006). The effect of type of written exercise on L2 vocabulary retention. *TESOL Quarterly, 40*(2), 273-293.

Hazrat, M. (2020). The Involvement Load Hypothesis and Its Impact on Vocabulary Learning (Doctoral dissertation, University of Auckland). Retrieved February 28, 2022, from https://researchspace.auckland.ac.nz/handle/2292/51729

Hulstijn, J. H. and Laufer, B. (2001). Some empirical evidence for the involvement load hypothesis in vocabulary acquisition. *Language Learning,* 51(3), 539-558.

Jones-Mensah, I., Tabiri, M. O., Fenyi, D. A., Kongo, A. E. and Amexo, D. (2022). Vocabulary knowledge of collocation in business texts: a case of ESL tertiary students. *International Journal of Education, Technology and Science(IJETS), 2*(1), 001–023.

Karalık, T. (2016) *The Effects of Task Induced Involvement Load Hypothesis on Turkish EFL Learners' Incidental Vocabulary Learning*. Unpublished master's thesis. Eskişehir: Anadolu Üniversitesi, Eğitim Bilimleri Enstitüsü.

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.

Kim, Y. (2008). The role of task-induced involvement and learner proficiency in L2 vocabulary acquisition. *Language Learning,* 58(2), 285-325.

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. and Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics*, 22(1), 1-26.

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

Mármol, G. A. and Sánchez-Lafuente, Á. A. (2013). The involvement load hypothesis: The effect on vocabulary learning in primary educaion. *Revista Española de Lingüística Aplicada*, (26), 11-24.

Pourakbari, A. A. and Biria, R. (2015). Efficacy of task-induced involvement in incidental lexical development of Iranian senior EFL students. *English Language Teaching,* 8(5), 122-131.

Sarani, A., Mousapour Negari, G. and Ghaviniat, M. (2013). The role of task type in L2 vocabulary acquisition: a case of involvement load hypothesis. *Acta Scientiarum. Language and Culture,* 35(4).

Sarbazi, M. R. (2014). Involvement load hypothesis: Recalling unfamiliar words meaning by adults across genders. *Procedia-Social and Behavioral Sciences,* 98, 1686-1692.

Teng, M. F., & Zhang, D. (2021). Task-induced involvement load, vocabulary learning in a foreign language, and their association with metacognition. Language Teaching Research, 13621688211008798. Retrieved February 28, 2022, from https://bit.ly/36SFJJ3

Walsh, M. I. (2009). *The involvement load hypothesis applied to high school learners in Japan: Measuring the effects of evaluation.* Unpublished master's thesis. United Kingdom: Birmingham University.

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.

Webb, S. A. (2009). The effects of pre-learning vocabulary on reading comprehension and writing. *Canadian Modern Language Review,* 65(3), 441- 470.

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

Yaqubi, B., Rayati, R. A. and Allemzade Gorgi, N. (2012). The involvement load hypothesis and vocabulary learning: The effect of task types and involvement index on L2 vocabulary acquisition. *Journal of Teaching Language Skills,* 29 (1), 145-163.

Zou, D. (2017). Vocabulary acquisition through cloze exercises, sentence-writing and composition-writing: Extending the evaluation component of the involvement load hypothesis. *Language Teaching Research*, 21(1), 54-75

ANADOLU ÜNİVERSİTESİ
SOSYAL VE BEŞERÎ BİLİMLER BİLİMSEL ARAŞTIRMA VE YAYIN ETİĞİ KURULU
KARAR BELGESİ

ÇALIŞMANIN TÜRÜ:	Yüksek Lisans Tez Çalışması
KONU:	Eğitim Bilimleri
BAŞLIK:	Görev Kaynaklı Katılım Yükü Hipotezinin Türk Yabancı Dil Öğrencilerinin Tesadüfî Kelime Öğrenimi Üzerine Etkileri (The Effects of Task Induced Involvement Load Hypothesis on Turkish EFL Learners' Incidental Vocabulary Learning)
PROJE/TEZ YÜRÜTÜCÜSÜ:	Dr. Öğr. Üyesi Gonca SUBAŞI
TEZ YAZARI:	Mehtap YORGANCI
ALT KOMİSYON GÖRÜŞÜ:	-
KARAR:	Olumlu

Prof.Dr. Coşkun BAYRAK
(Başkan-Eğitim Fak.)

Prof.Dr. T. Volkan YÜZER
(Başkan Yardımcısı-Açıköğretim Fak.)

KATILMADI

Prof.Dr. Esra CEYHAN
(Eğitim Fak.)

Prof.Dr. Münevver ÇAKI
(Güzel Sanatlar Fak.)

Prof.Dr. M. Erkan ÜYÜMEZ
(İkt. ve İdari Bil. Fak.)

Prof.Dr. Handan DEVECİ
(Eğitim Fak.)

Prof.Dr. Emel ŞIKLAR
(İkt. ve İdari Bil. Fak.)