

## An empirical research on the adequacy of task-induced involvement load hypothesis: A case for intentional vocabulary learning

Ali Jahangard\*, Sharif University of Technology, Sharif Languages and Linguistics Center, Tehran, Iran

### Suggested Citation:

Jahangard, A. (2021). An empirical research on the adequacy of task-induced involvement load hypothesis: A case for intentional vocabulary learning. *International Journal of Learning and Teaching*. 13(2), 94-105. <https://doi.org/10.18844/ijlt.v13i2.5800>

Received November 23, 2020; revised February 15, 2021; accepted April 16, 2021.

Selection and peer review under responsibility of Prof. Dr. Hafize Keser, Ankara University, Ankara, Turkey.

©2021 Birlesik Dunya Yenilik Arastırma ve Yayıncılık Merkezi. All rights reserved.

### Abstract

The present study aimed at examining the adequacy of the task-induced involvement load hypothesis in intentional learning. An investigation was carried out to find out whether proficiency level of learners had a role in the effectiveness of the vocabulary tasks with different involvement loads. One hundred and thirty-six university students were randomly assigned into four task groups, each of which included upper and lower intermediate learners. Reading comprehension and discussion, reading comprehension and gap filling, reading comprehension plus sentence-making and reading comprehension plus translation with different involvement loads were compared against each other in terms of the immediate and delayed retention of new words. The study partially supported the involvement load hypothesis in that the task with the highest involvement loads resulted in better immediate and delayed retention of new words. The results of the experiment also showed that tasks with similar involvement loads might not result in similar amounts of vocabulary learning.

**Keywords:** Task-induced involvement, load hypothesis, vocabulary learning, word retention, task.

---

\* ADDRESS FOR CORRESPONDENCE: Ali Jahangard, Sharif University of Technology, Tehran, Iran.  
E-mail address: [jahangard@sharif.edu](mailto:jahangard@sharif.edu)

## 1. Introduction

One of the overwhelming hurdles in learning a foreign or second language is to learn a huge number of words in a comparatively short period of time, a challenge most of learners have to confront in their efforts to learn a second or foreign language. Supporting learners in coping with the learning of thousands of new words obliges teachers to explore and discern the most effective tasks for vocabulary learning. Many teaching practitioners and researchers have been cognizant of this requisite, but most often they have been in a state of indeterminacy regarding what tasks are the most effective in helping learners deal with the problem. Many language teachers and researchers have felt that some tasks are more effective than others in helping the learners retain the meaning of words in the course of practical experience as either a second language learner or teacher or both. There appears to exist a general consensus among vocabulary acquisition researchers that there is a positive correlation between the amount of mental effort which is invested on the learning of target words and the retention of those words. However, converting the extent of mental effort into an index which is operationally amenable to measurement has been a wild horse which has stubbornly defied the attempts of the applied linguists to bridle.

In this connection, Laufer and Hulstijn (2001, p. 17) proposed the task-induced involvement load hypothesis (ILH) which claimed that 'tasks with a higher involvement load will be more effective for vocabulary retention than tasks with a lower involvement load'. Later, they came up with some criteria for assessing the involvement load in their hypothesis. Likewise, Craik and Lockhart (1972), many years earlier, proposed the 'depth of processing' which professed that the storage of new information in the long-term memory is not determined by the amount of time dedicated to it in short-term memory, rather it is the depth of its processing which holds sway on the success of the information retrieval. Building on this hypothesis, subsequently, Laufer and Hulstijn (2001), in an attempt to assuage the shortcomings of the depth of processing and redeem it into an operationally implementable theory, proposed the task-induced ILH which dissects any given task into three constituent components: need, search and evaluation, each of whose components could be present or absent at a time. According to Laufer and Hulstijn (2001), the combination of these three factors indicates the amount of involvement load induced by a task which, in turn, can decide the retrieval of the given unknown words embedded in it in later uses. While some other studies have addressed the hypothesis in an incidental learning setting, there is a void to be filled when it comes to the evaluation of the theory in an intentional setting. Therefore, this study sets out to test the adequacy of the hypothesis in an intentional learning setting.

### 1.1. Involvement load hypothesis

Taking into consideration the depth of processing and elaboration concepts, Laufer and Hulstijn (2001) claimed that these concepts are difficult to formulise and they lacked an operational definition. To create a change, they suggested a construct to operationalise the concepts 'attention' and 'elaboration' into concrete task-specific constructs. To this end, they proposed the construct of *involvement*. Involvement was defined as 'a motivational-cognitive construct which can explain and predict learners' success in the retention of hitherto unfamiliar words' (ibid, p. 14). By the term 'cognitive' in the above-mentioned definition, they meant the information processing involved in the decoding of unknown words. They assumed that 'retention of words when processed incidentally is conditional upon three factors in a task: need, search, and evaluation'. Laufer and Hulstijn (2001) introduced the *involvement* construct and considered need, search and evaluation as the components of involvement and stated that each of these three components can be present or absent in an activity for vocabulary learning. This new construct which according to Laufer and Hulstijn (2001), is a start to operationalise the traditional concepts such as 'attention', 'noticing', 'elaboration' and 'motivation' into task-specific components, decomposes the concept of processing into two categories: cognitive and motivational components.

According to Laufer and Hulstijn (2001), *need* is a motivational and non-cognitive component of involvement. It has two dimensions: moderate and strong. Need is moderate when it is imposed by an external agent (e.g. when the teacher asks learners to fill in the blanks). On the other hand, it is strong when learners themselves want to do something (e.g. when learners decide to find the meaning of a new word). *Search* and *evaluation* are the two cognitive components of involvement. Laufer and Hulstijn (2001) state that these two factors, concerned with information processing, are related to noticing, proposed by Schmidt (1994a, 1994b, 2000), and attention. *Search* is defined as the learners' attempt to find the meaning of an unknown word. This attempt can be consulting a dictionary or a teacher in the classroom. *Search* is categorised as present or absent. *Evaluation* is defined as the multilateral comparison and contrast of the words with each other and assessing their appositeness for a particular context. Laufer and Hulstijn (2001, p. 15) contented that '*evaluation* implies some kind of selective decision based on a criterion of semantic and formal appropriateness (fit) of the word and its context'. For them, *evaluation* fell into two categories: moderate and strong. *Evaluation* is moderate if a comparison is made between some words in order to fit a specific context. A strong evaluation demands 'making a decision about additional words which will combine with the new word in an original sentence or text' (ibid, p. 15).

Based on the components of involvement, Laufer and Hulstijn (2001, p. 15, 20) proposed ILH. Involvement load is defined as the 'combination of the presence or absence of the involvement factors *need*, *search* and *evaluation*'. According to the proposed theory, they made the postulation that 'other factors being equal, words which are processed with higher involvement load will be retained better than words which are processed with lower involvement load'. They pointed out that their proposal was 'to explore the relationship between retention and various aspects of deep processing'. Consequently, they proposed the *involvement index* to operationalise the ILH. The involvement index was a numerical factor showing the number of presences or absences of each component. For the absence of each involvement component, the component was marked as zero and the presence of it was marked as one. Moreover, a moderate index was marked as one and a strong index as two. Accordingly, the degree of the involvement load of a task or activity for learning a new word was practical and measurable. Laufer and Hulstijn (2001, p. 20), however, did not stop at this point and claimed that 'although involvement will vary with words, involvement can also be influenced by task factors' with different degrees of *need*, *search* and *evaluation*.

There are few empirical studies which have directly investigated the adequacy of the involvement load hypothesis (e.g. Hulstijn & Laufer, 2001; Keating, 2008; Kim, 2008; Jahangard, 2013; Jahangard & Akbari, 2013; Jing & Jianbin, 2009; Yagubi, Rayati, & Gorgi, 2010). In sum, the majority of the studies mentioned above, although with mixed results in some cases, suggest that the ILH has a considerable predictive power in prognosticating the degree of efficacy of the learning tasks in incidental vocabulary learning contexts. However, no studies to date, to the best of our knowledge, have been conducted to probe into the predictive adequacy of the ILH in intentional vocabulary learning contexts. The question of whether the theory can adequately predict the retention of words based on the index of involvement load induced by the related tasks is still to be answered. At this point, it may be necessary to elaborate on incidental and intentional learning and the differences between these two. Essentially, in incidental learning, the attention of the learner is primarily focused on the comprehension of a given text rather than on the formal features of the passage, such as the lexical items and grammatical forms in it (Schmidt, 1994a). However, in intentional vocabulary learning contexts, the learner is already aware that that main purpose of the reading and the related tasks is to learn the target vocabulary and consequently his/her attention is consciously concentrated on the target words and, for this very reason, it is very likely that the learner uses other cognitive strategies, such as repetition, mnemonics, keywords, translation etc. to boost learning of the lexical items. These additional cognitive efforts which are not under the control of the instructor might change the balance of the controlled involvement load of the designed tasks. Hence, the pre-planned involvement load index of each one of the designed tasks might change and lose its functionality in the prediction of the amount of learning that might take place. Nevertheless, this is simply hypothesising and the

relationship of involvement load and retention of words in an intentional setting needs to be further studied. Moreover, the necessity for the study becomes more evident in light of the contention that intentional learning is one of the most effective and common routes through which learners attempt to learn new lexical items, particularly in the EFL context where exposure to L2 is very limited. Schmitt (2008, p. 342) asserts that, 'although research has demonstrated that valuable learning can accrue from incidental exposure, intentional vocabulary learning almost always leads to *greater and faster gains, with a better chance of retention and of reaching productive levels of mastery*' (emphasis is ours). Different definitions of intentional learning have been proposed from among which we subscribe to the methodological definition of the term. In this view, intentional learning is defined as performing a task without being told in advance that the learners will be tested afterwards (see Schmitt, 2008).

## **2. Purpose of the study**

Building on the ILH proposed by Hulstijn and Laufer (2001), this study aimed to investigate the effectiveness of the involvement load in the vocabulary learning tasks across two proficiency levels in an 'intentional' setting. The predictive adequacy of the task-induced ILH was investigated through conducting experiments. The study aimed to test the effectiveness of tasks with different levels of involvement across two proficiency levels in terms of immediate and delayed retention of new words. To do so, three different vocabulary learning tasks with varying levels of involvement were designed and administered, i.e. reading comprehension plus discussion (involvement load = 1), reading comprehension and gap filling (involvement load = 2), reading comprehension plus sentence-making (involvement load = 3) and reading comprehension plus backward translation (involvement load = 3). An additional purpose in the current research was to compare the effectiveness of two tasks with the similar involvement loads.

## **3. Research questions**

The study was conducted to answer the following research questions:

1. Does the level of task-induced involvement affect the immediate retention of new words by EFL learners from two different proficiency levels when tasks with different levels of involvement are administered?
2. Does the level of task-induced involvement affect the delayed retention of new vocabulary by EFL students at two different proficiency levels when tasks with different levels of involvement are administered?
3. Does task type affect the amount of the immediate retention of new words by EFL students at two different proficiency levels when two tasks with the same level of involvement are administered?
4. Does task type affect the amount of delayed retention of new words by EFL students at two different proficiency levels when two tasks with the same level of involvement are administered?

## **4. Methodology**

### **4.1. Participants**

A sample of 136 male and female university students, with an average age of 19, participated in the study. The participants were undergraduate engineering students at Sharif University of Technology in Iran. Their first language was Persian and all were taking a general English course as part of their curriculum.

A sample TOEFL which was developed by the researcher, in accordance with the characteristics of the TOEFL designed by the ETS, was used to classify the participants into upper intermediate (UI) and lower intermediate (LI) levels. The validity of the sample TOEFL named 'best properties' was

corroborated by the experts of Sharif University of Technology. Cronbach's alpha reliability of the scores generated by the test was 0.92, which shows that the internal consistency of the test was quite acceptable. The test consisted of 100 multiple-choice questions, making up two sections: section one included 40 questions concerning structure and section two included 60 questions concerning vocabulary and reading comprehension. The participants were assigned into UI and LI levels based on their TOEFL scores. One standard deviation above the mean (above 50) and one standard deviation below the mean (30–50) were considered as UI and LI, respectively.

#### **4.2. Instrumentation**

Instrumentation lent itself to four components: target vocabulary, a reading passage, tasks and vocabulary tests, each of which will be dealt with in the following sections.

##### **4.2.1. Target vocabulary**

A total of 10 target words were chosen according to the following procedure: first, a list of 56 words was prepared for the students and they were asked to provide the equivalent Persian (L1) or English meaning of the words which they knew. By doing so, the known words to the participants were discarded and the 10 target words to which none of the participants had provided correct answers and, therefore, were unfamiliar to them were chosen for the study. The target words consisted of five adjectives: three nouns, one verb and one adverb: *stunted, mercenary, callous, perverse, subservient, cynic, propensity, pariah, condone and aloof*.

##### **4.2.2. Reading passage**

The 10 unfamiliar words chosen by the participants as the target words were embedded in a passage. The passage, consisting of 408 words, was based on four articles taken from the book *1,100 Words*. Factors such as length, syntactic complexity of the constituent sentences, test-difficulty level and topic familiarity of the text were considered through keeping the grammar of the text simple and inserting textual clues into the text. The appropriacy of the text was confirmed by three experienced teachers. The passage was followed by seven comprehension questions whose correct answers were contingent on knowing the target vocabulary.

##### **4.2.3. Vocabulary tests**

Two vocabulary tests were used in this study: an immediate post-test (immediately after the treatments) and a delayed post-test (1 week after the treatment).

Immediate and delayed post-tests included 10 unknown words in the form of a multiple-choice test which was administered to the participants. The two vocabulary post-tests had the same format and content, but the order of the items and the alternatives was different. The post-tests consisted of 10 items, each of which included 1 underlined word with 4 choices. The participants were required to select the alternatives closest in meaning to the underlined words. Also, the reliability indexes (Cronbach's  $\alpha$ ) of the immediate post-test and the delayed post-test were 0.82 and 0.88, respectively.

#### **4.3. Procedure**

##### **4.3.1. Design**

All the participants were from 8 intact classes, each of which included 24 students on average. The participants were randomly assigned into four task-type groups. In other words, each two classes were assigned into one specific task. They were also grouped into two levels of proficiency UI and LI based on their paper-based TOEFL scores (UI: above 50; LI: 20–50). The schematic design of the experiments of the study is illustrated in Table 1.

**Table 1.**

<b>Treatment</b>	<b>Task group 1</b>	<b>Task group 2</b>	<b>Task group 3</b>	<b>Task group 4v</b>
Treatment	Reading and discussion	Reading and sentence-making	Reading and translation	Reading and involvement
Load	Need = 1	Need = 1	Need = 1	Need = 1
Index	Search = 0	Search = 0	Search = 0	Search = 0
Evaluation = 0	Evaluation = 1	Evaluation = 1	Evaluation = 2	Evaluation = 2
Time on task	40 m	40 m	40 m	40 m
Sample size	30	30	40	36
Proficiency Level	UI = 10 LI = 20	UI = 14 LI = 16	UI = 16 LI = 24	UI = 15 LI = 21

#### **4.3.2. The design of the experiment**

The data for the experiments were collected during the participants' class time. It took three sessions to gather the data for each task group, i.e. one session for TOEFL, one for treatment and immediate vocabulary test administered on the same day and one session for the delayed vocabulary test. The treatment session and the other sessions were carried out by the researchers. For the treatment, the participants were randomly assigned into four task-type groups, i.e. reading comprehension plus discussion (involvement load = 1), reading and filling in the blanks (involvement load = 2), reading plus making sentences (involvement load = 3) and reading plus translation (involvement load = 3). The participants in each task group were categorised into two proficiency levels, UI and LI, based on their paper-based TOEFL scores (UI: one standard deviation above the mean; LI: one standard deviation below the mean). Time on task which was controlled in all the four tasks was 40 minutes. The participants were asked to provide the meaning of the words which they knew from a list of 56 words. Based on the words unfamiliar to the participants, the target vocabulary was selected and the reading passage was constructed.

Each task-type group went through four steps: 1. pre-reading; 2. Reading; 3. post-reading, including reading comprehension questions, a task; and 4. a vocabulary post-test. The four steps were the same for all the groups except for step 3, which was different for each task group. In group one, the participants were asked to have a discussion as an activity. For task two, they had a fill-in-the-blanks activity and task three involved making sentences in English using the target words. In group four, the participants were required to translate the Persian sentences into English, and they were told that using the target vocabulary in the sentences was obligatory. It is worth noting that the participants were asked to follow all the steps in their scheduled time set for each step. Therefore, the same procedure was followed step by step for each group.

Before the treatment, the participants in each task group were given papers including a three-page article along with an answer sheet. The steps for each task group were introduced and the steps to be followed by the participants were written on the board in English by the researchers. All the participants in each class were asked to participate in activities and follow the procedures step by step during the class time. They were informed of the incoming vocabulary test which was to be given after the treatment.

The first step in each task group was a pre-reading which included a brainstorming on the topic of the text. In this step, the researchers of the study wrote some words related to the text on the board and asked the learners to think about the words and share their ideas with their classmates. This step, which took 5 minutes across the classes, was implemented to help participants activate the related schemata in their minds for a better comprehension of the upcoming reading text. Also, the learners were free to choose either their L1 or English to communicate their ideas on their own volition.

The second and third steps required participants to read the text silently and answer the seven comprehension questions that ensued. They were told to use the marginal glossary to find the meaning of the unknown words. The researchers helped them with any other difficulty in the text.



Then, as the fourth step, they did the relevant task for each task-type group as stated above. Immediately, the immediate vocabulary test was given, which took 8 minutes of the class time. The delayed vocabulary post-test was given a week later. However, contrary to the immediate post-test, the participants were not informed of the upcoming delayed post-test that was taken 1 week later. The reason for keeping the participant uninformed of the delayed post-test was that the researchers' primary intention was to assess the consistency of the learning which had occurred by the aid of the tasks with controlled and pre-determined variations in the amount of involvement loads, over time. If the participants were aware of the upcoming delayed post-test, it was quite likely that they would do some other extra practices on the target words in the interim time to refresh their minds and to perform better in the delayed post-test, an act which would evidently contaminate the test results.

#### **4.3.3. Tasks**

Based on Laufer and Hulstijn's (2001) hypothesis, four learning tasks with variant levels of involvement were used in the study (see Table 1): reading and discussion (involvement load = 1), reading plus filling in the blanks (involvement load = 2), reading and sentence-making (involvement load = 3) and reading plus translation (involvement load = 3). In fact, the reading passage was identical across the tasks and what varied from one task to other was the way word meanings were presented to the participants and also the type of practice they were supposed to do with the given lexical items.

The first three tasks were different in terms of involvement load, and tasks 3 and 4 had the same involvement load. Participants in task group one received a reading passage with marginal glosses; they had to read and answer the reading comprehension questions accordingly. The list of words glossed in the margin included 10 unknown words highlighted in bold print in the text along with their equivalent Persian, as well as English meaning. After they read the text and answered the relevant comprehension questions, participants were asked to have a discussion on the topic of the passage. This extra activity was added to control the equal 'time on task' across all four tasks. However, some of the participants used the unknown words as they were sharing the ideas. In this task, the involvement load index was one. In other words, the need, search and evaluation components, according to Hulstijn and Laufer (2001), had an index of 1, 0 and 0, respectively. The need was moderate (i.e. 1) because the need to use the glossed list was imposed by the task itself not the participants. Search and evaluation had an index of 0 because there was no searching for the meaning and no comparing of the words.

Group two included a reading and fill-in-the-blanks task in which participants were asked to read the same passage with the same list of words glossed in the margin and answer the seven comprehension questions. Moreover, they participated in a gap-filling task (GT). A list of 10 statements including 10 blanks was given to the participants. They were asked to read the statements and fill in the blanks using the words in the list (involvement load = 2). The involvement index of need was one because it was moderate; the meaning of the unknown words was already provided, and therefore the search index was one; the words in the list needed to be evaluated against each other and hence there was a comparison among unknown words and the appropriateness of the context, and therefore the evaluation index was one.

Task group three offered a reading task plus writing sentences, in which the participants read the passage with some words glossed in the margin and answered the same comprehension questions. Moreover, they were required to make sentences using the words in the list (involvement load = 3). In this task, the participants were told to use the glossed words in the margin to get the meaning of the unknown words. Therefore, the need was moderate with an index of 1. Search was absent (i.e. 0) because the words and their corresponding meanings were given in the list of glossed words, and evaluation was strong and equalled two because the participants had to compare and use the appropriate words in new contexts.

Finally, task group four was a reading plus translation task (TT) in which the participants needed to read the same passage with glossed words in the margin, answer the comprehension questions and

translate 10 sentences written in Persian including the unknown glossed words into English. Involvement load in this task was 3 [based on Laufer and Girsai, (2008)]. This task was the same as task three according to the Hulstijn and Laufer's (2001) hypothesis. That is, the need, search and evaluation components had an index of 1, 0 and 2, respectively. The evaluation index was two because according to Laufer and Girsai (2008), in L1–L2 translation, the target word is evaluated against the other words surrounding it and also the new context is created by the learner.

## 5. Results

The descriptive statistics of the immediate and delayed vocabulary post-tests of both UI and LI proficiency levels and also the total means which are the mean scores of each one of the groups across from each one of the tasks are displayed in Table 2.

**Table 2.**

Immediate		Delayed		
	N	Mean	N	Mean
UI				
Discussion	10	8.2	10	7.6
Results of descriptive statistics for immediate and delayed post-tests				
Gap-filling	14	8.4	14	7.0
Sentence-making	16	9.3	16	9.0
Translation	15	8.7	15	8.0
LI				
Discussion	20	6.8	20	5.2
Gap-filling	16	7.5	16	6.0
Sentence-making	24	8.3	24	7.9
Translation	21	8.9	21	5.9
Total				
Discussion	30	7.5	30	6.0
Gap-filling	30	7.9	30	6.4
Sentence-making	40	8.8	40	8.3
Translation	36	8.3	36	6.8

The descriptive statistics showed that the participants in the sentence-making task (SMT) group had the best performance at both proficiency levels and gained the highest mean scores among the three task groups on both vocabulary post-tests (Table 2).

The first research question asked whether the level of task-induced involvement affects immediate retention of words by EFL learners from two different proficiency levels when three tasks with different levels of involvement are administered. The two-way ANOVA findings of the immediate vocabulary post-test revealed that there was no significant interaction between the proficiency level and the level of involvement,  $F(3, 128) = 0.380, p = 0.768, \eta^2 = 0.009$ . There was indeed a significant main effect for task,  $F(3, 128) = 8.243, p = 0.058, \eta^2 = 0.892$ . The results also showed that on the immediate vocabulary post-test, there was a significant main effect for proficiency level,  $F(1, 128) = 21.289, p = 0.016, \eta^2 = 0.869$ . The Bonferroni *post-hoc* comparisons indicated that only the mean score of the SMT group ( $M = 8.8$ ) was significantly higher than that of the discussion task (DT) group ( $M = 7.5$ ), and other comparisons were not significant.

The second research question investigated whether the level of task-induced involvement affects the amount of delayed retention of new vocabulary by EFL students at two different proficiency levels when three tasks with different levels of involvement are administered. The results of two-way ANOVA showed no significant interaction between the involvement load and the proficiency level on the delayed post-test,  $F(3, 128) = 0.888, p = 0.449, \eta^2 = 0.02$ . There was a significant main effect for



task,  $F(3, 128) = 8.100, p = 0.060, \eta^2 = 0.89$ . Moreover, there was a significant main effect for proficiency level,  $F(1, 128) = 23.206, p = 0.016, \eta^2 = 0.882$ .

The Bonferroni *post-hoc* comparisons indicated that on the delayed post-test, the mean score of the SMT group ( $M = 8.37$ ) was significantly higher than that of the gap-filling ( $M = 6.46$ ) and the discussion ( $M = 6.03$ ) task groups, and the difference between the discussion and the GT groups was not significant. The mean score of the two proficiency levels also was found to be significantly different on the delayed post-test. That is, the participants at UI level received a higher mean score ( $M = 7.888$ ) than those at LI level ( $M = 6.389$ ) in the delayed post-test across the three task groups.

The concern of the third question was whether there was an effect of task type on the amount immediate word retention by EFL learners when two tasks with the same level of involvement were administered. The Bonferroni *post-hoc* comparisons indicated that the difference between the SMT group ( $M = 8.72$ ) and the TT group ( $M = 8.36$ ) was not significant on the immediate vocabulary post-test ( $p > 0.05$ ).

The fourth research question asked whether there was an effect of task type on the amount of the delayed retention of new vocabulary by EFL learners when two tasks with the same level of involvement were administered. The findings indicated that the mean score of the SMT group and the TT group in delayed vocabulary post-test was significantly different in favour of the SMT group ( $M = 8.4$ ) over the TT group ( $M = 6.9$ ),  $p > 0.05$ .

## 6. Discussion

To investigate the adequacy of the ILH (Laufer & Hulstijn, 2001) for intentional learning, an experiment was conducted in the present study.

In sum, the results of the immediate retention were as follows: SMT: 8.8 > TT: 8.3 > GT: 7.9 > DT: 7.5. Also, the involvement load indices assigned to each one of the tasks ranked in the following order: SMT: 3 = TT: 3 > GT: 2 > DT: 1.

Analysis of the means from TT and SMT showed that the difference was not statistically significant, which is consistent with the prediction made by the ILH. Likewise, congruent with the prediction of the ILH, the difference between the means from the SMT and the DT proved significant. However, the results showed no significant differences between the means of the TT, GT and DT. This means that the ILH has not been successful in predicting the results from these tasks. So far, one can conclude that the results of the study in the immediate retention of words only partially support the involvement theory, a claim which is consistent with the findings of similar studies (e.g. Hulstijn & Laufer, 2001; Jahangard, 2013; Jahangard & Akbari, 2013; Jing & Jianbin, 2009; Kim, 2008).

In the delayed retention of words, the means from the different tasks ranked in the following order: SMT: 8.3 > TT: 6.8 > GT: 6.4 > DT: 6.

Only the difference between the mean from SMT proved to be statistically different from the other tasks in the study and other comparisons of the means were not significant. This probably means that the actual involvement load induced by the SMT is greater than others – although theoretically equal with the TT – and consequently, it has led to a relatively more durable learning. Regarding the significant difference between the means of the SMT and TT, which looks to be incompatible with the predictions made by the ILH, a possible explanation might be that the quality and depth of the involvement induced by the SMT have been slightly different from those of the TT. These minuscule differences might have caused fleeting dissimilarities, although invisible in the immediate retention, and have lingered to an influential effect on the delayed retention. In fact, the minute intangible differences in the quality of the involvement load induced by the SMT might have attenuated the attrition rate of learning ensued from the task. In the SMT, the learners had to map the intended meaning into an appropriate syntactic structure in accordance with the syntactic properties of the related individual lexical components. However, in the TT, the mapping of the intended meanings had

already been forged into a syntactic structure in the source language. In many cases, this syntactic structure in the source language might be very similar to the one needed in the target language; thereby, the need to generate the related syntactic structure(s) in the target language by the learner becomes obviated. This can affect the amount of mental effort which is invested on completing the TT. Therefore, one can argue that the quality, amount and depth of the mental effort induced by the TT might have been different from the SMT in the experiment.

A further explanation on the inconsistency of the results from the DT, GT and TT with the predictions made by the involvement load theory might be attributed to the fact that the theory in essence was proposed to explain incidental learning, whereas in the current study, the intentional learning of the words was examined. It is likely that the learners resorted to other cognitive strategies such as repetition, mnemonics keywords, note-taking, reviewing etc. rather than comparison and selection processes embedded in the evaluation component as characterised by the theory. The replacement of the more involvement inducing mental procedures with the more effort-economical conscious processes might have neutralised the intended involvement loads of the given tasks.

Also, the apparent inconsistency can be due to inappropriate assigning of the involvement indices to the tasks at issue. In other words, the indices assigned to each one of the tasks regarding the involvement load components are needed to be re-tuned. This argument becomes more convincing if the attrition rate of the learning ensued from each one of the tasks is closely examined. The attrition rates are as follows:

DT = -1.5, GT = -1.5, SMT = -0.5, and TT = -1.5.

As it is evident, the attrition rate of retention from the different tasks is equal, except for SMT. This probably indicates that because SMT induces a greater involvement load, the least attrition has occurred as a consequence. If this re-tuning of the indices is applied to the immediate retention involvement load assigning, the results would become more consistent with the predictions of the theory except for the TT.

The experiments in the present study also investigated the role of proficiency level on the effectiveness of the tasks with different involvement indices. The results revealed that more proficient participants across the task groups performed relatively better than the less proficient ones on both immediate and delayed post-tests. The findings of the present study do not lend support to Kim's (2008) study which found no significant difference between learners' retention of new words in incidental learning across the two different proficiency levels. One possible explanation is that knowing more features of a new word will increase the possibility of learning that word. According to 'lexical quality' hypothesis, to become more proficient in lexical quality, an effort is needed to acquire all the features of a lexicon, i.e. orthography, phonology, grammar and meaning (Perfetti, 2007). Moreover, 'noticing hypothesis' claims that to change input into intake for learning, learners need to notice forms and their related meanings. Therefore, the incongruity of the findings of the current study and those of Kim (2008) could be rooted in the fact that more proficient learners, presumably, might have had a more elaborate network of vocabulary knowledge enabling them to notice additional features of the target words, such as collocations, syntactic category and semantic nuances, which distinguishes a lexical item from the synonyms which can be associated with a given word. Therefore, it can be argued that the quality of the knowledge of the target words which was stored in the mental lexicon of the more proficient learners might have been higher than that of the less proficient learners, which, in turn, could culminate into better retention rates by the more proficient learners.

## 7. Conclusion

The results of the study appear to demonstrate that the task-induced involvement hypothesis falls short of predictive adequacy in envisaging the learners' vocabulary retention as far as intentional learning is concerned. However, this by no means devalues the substantial position of the theory in

incidental vocabulary learning. Moreover, it is noteworthy that the findings of the current study point to a partial harmony between the involvement load and the retention rate in higher proficiency levels. However, it seems that for the hypothesis to become more reliable regarding its predictive power for intentional vocabulary learning, more refinements and tunings of the hypothesis seem inevitably necessary.

Unlike the findings of the study by Kim (2008), which showed no significant effect of language proficiency on incidental vocabulary learning, the current study showed that proficiency level played an important role in escalating the retention rate of new words in intentional learning.

Interestingly, it appears that the involvement load which is induced by sentence production tasks influences the retention rate more strongly than other presumably parallel tasks of the same involvement loads.

In addition, the evidence from the present study provides some implications for L2 pedagogy in the case of vocabulary acquisition. First and foremost, teachers could employ different tasks with high involvement loads in their teaching of vocabulary particularly to high proficiency learners to enhance intentional learning.

Secondly, it is suggested that teachers employ sentence-production tasks in their vocabulary teaching processes. The findings of the present study speak to the advantage of sentence-making activities for better retention of words.

Interestingly enough, in light of the findings of the present study, backward translation proved an efficient task, very close to its counterpart, i.e. sentence production, in terms of the retention rates that were obtained from the tasks. Hence, teachers are well advised to include L1–L2 TTs in their vocabulary teaching.

## References

- Craik, F. I. M. & Lockhart, R. S. (1972). Levels of processing: a framework for memory research. *Journal of Verbal Learning and Verbal Behaviour*, 11, 671–684.
- Hulstijn, J. H. & Laufer, B. (2001). Some empirical evidence for the involvement load hypothesis in vocabulary acquisition. *Language Learning*, 51, 539–558.
- Hyde, T. S. & Jenkins, J. J. (1969). Differential effects of incidental tasks on the organization of recall of a list of highly associated words. *Journal of Experimental Psychology*, 82, 472–491.
- Jahangard, A. & Akbari, S. (2013). Form-focused EFL vocabulary learning: a case for translation. *Iranian Journal of Language Issues*, 1(1).
- Jahangard, A. (2013). Task-induced involvement in L2 vocabulary learning: a case for listening comprehension. *Journal of English Language Teaching and Learning*, 5(12), 43–62.
- 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. 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. & Hulstijn, J. H. (2001). Incidental vocabulary acquisition in a second language: the construct of task-induced involvement. *Applied Linguistics*, 22, 1–26.

- Jahangard, A. (2021). An empirical research on the adequacy of task-induced involvement load hypothesis: A case for intentional vocabulary learning. *International Journal of Learning and Teaching*, 13(2), 94-105. <https://doi.org/10.18844/ijlt.v13i2.5800>
- Laufer, B. & Girsai, N. (2008). Form-focused instruction in second language vocabulary learning: a case for contrastive analysis and translation. *Applied Linguistics*, 29, 694–716.
- Perfetti, C. (2007). Reading ability: lexical quality to comprehension. *Scientific Studies of Reading*, 1, 357–383.
- Schmidt, R. (1994a). Deconstructing consciousness in search of useful definitions for applied linguistics. In J. H. Hulstijn & R. Schmidt (Eds.), *Consciousness in second language learning* (vol. 11, pp. 11–26). AILA Review.
- Schmidt, R. (1994b). Implicit learning and the cognitive unconscious: of artificial grammars and SLA. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 165–210). London, UK: Academic Press.
- Schmidt, N. (2000). *Vocabulary in language teaching*. Cambridge, UK: Cambridge University Press.
- Schmitt, N. (2008). Review article: instructed second language vocabulary learning. *Language Teaching Research*, 12(3), 329–363.
- Yagubi, B., Rayati, R. A. & Allemzade Gorgi, N. (2010). The involvement load hypothesis and vocabulary learning: the effect of task types and involvement index on L2 vocabulary acquisition. *The Journal of Teaching Language Skills (JTLS)*, 2(1), 145–163.