Porta Linguarum 33, enero 2020 265-278

Vocabulary Learning And Retention: Cognitive Load Framework On Trial

Abdolvahed Zarifi
Zeinab Azizinezhad
Yasouj University, Iran

Received: 18 October 2019 / Accepted: 30 Marcha 2020
ISSN paper edition: 1697-7467, ISSN digital edition: 2695-8244

ABSTRACT: The literature has witnessed a large number of studies investigating the merits and effectiveness of the available methods of vocabulary teaching and learning in different EFL/ESL contexts among which one can refer to the involvement load hypothesis (ILH). Despite its widespread use, some criticisms have been levelled against this model. A new framework, namely cognitive load framework (CLF), has recently been proposed. Although CLF has been validated by some TESL experts, it has not, yet, been put to the acidity test. Therefore, the present quasi-experimental study was carried out to determine whether activities with higher cognitive load degrees as predicted by CLF framework are more effective for vocabulary learning and retention. To this end, a sample of 60 Iranian EFL learners was assigned to three experimental groups and was exposed to vocabulary learning tasks with different cognitive load indices for eight weeks. Experimental group 1 received the tasks with high cognitive load, while experimental groups 2 and 3 received the medium and low cognitive load tasks, respectively. The findings revealed that the vocabulary tasks with the highest cognitive load were the most effective, and those with the lowest load were the least effective in vocabulary learning and retention.
Key words: cognitive load framework, EFL learning, involvement load hypothesis, vocabulary learning, vocabulary retention.

Aprendizaje del bulario y Retención: Marco de Carga Cognitiva de Prueba.

RESUMEN: La literatura ha tenido en cuenta un gran número de estudios que investigaron los méritos y la eficacia de los métodos disponibles de enseñanza y aprendizaje de vocabulario en diferentes contextos EFL / ESL entre los cuales podemos incluir la hipótesis de carga de participación. A pesar de su uso general, algunas críticas han sido niveladas contra este modelo. Se ha propuesto recientemente un nuevo marco, marcado de carga normalmente cognitivo (CLF). Aunque CLF ha sido validado por algunos expertos en TESL, aún no se ha puesto en la prueba de acidez. El presente estudio de diseño cuasi-experimental pretest-posttest se llevó a cabo para determinar si las actividades con un alto grado de carga cognitiva, según lo predicado por el marco CLF son más efectivas para el aprendizaje y la retención de vocabulario. Para este fin, una muestra de 60 estudiantes de EFL Iraníes fue asignada a tres grupos experimentales y fue expuesta a las tareas de aprendizaje de vocabulario con diferentes índices de carga cognitiva durante 8 semanas. El grupo experimental 1 recibió las tareas con alta carga cognitiva, mientras que los grupos experimentales 2 y 3 recibieron las tareas de carga cognitiva media y baja, respectivamente. Los resultados revelaron que las tareas de vocabulario con la carga cognitiva más alta fueron las más efectivas, y las que tenían la carga más baja fueron las menos efectivas en el aprendizaje y la retención de vocabulario.
Palabras clave: aprendizaje de vocabulario, aprendizaje EFL, hipótesis de carga de implicación, marco de carga cognitiva, retención de vocabulario.

1. Introduction

Nowadays, vocabulary learning is believed to be at the heart of language learning for both teachers and learners. In fact, learning vocabulary is an essential part of acquiring a second language and learners with higher vocabulary knowledge tend to perform better in written or oral learning tasks. Among others, Richards and Renandya (2002) maintain that vocabulary is "a core component of language proficiency and provides much of the basis for how well learners speak, listen, read, and write" (p. 255). In other words, vocabulary knowledge strongly affects learners' success in other language skills, hence the significance of studying efficient ways of teaching and learning vocabulary.

Due to the importance of vocabulary learning and teaching, Lewis (1993) argues, "language consists of grammaticalized lexis, not lexicalized grammar" (p. 34). Likewise, vocabulary items are considered as building blocks of any language learning without which very little or no messages can be conveyed. Therefore, achieving an acceptable degree of proficiency in any language is threatened by the daunting task of mastering a huge bulk of new words.

Regarding the vocabulary learning, different views have been put forward. Some consider second language vocabulary learning as an incidental process (Krashen, 2008; Pellicer-Sánchez & Schmitt, 2010; Stahl & Nagy, 2006); however, other researchers have claimed that incidental learning by itself could not bring about efficient vocabulary learning and retention. In addition, it has been argued that intermediate and advanced L2 learners rarely look up the meanings of unknown words in order to understand the main idea of a text (Hulstijn, 1992). Learners often ignore unfamiliar words except when the words are essential for understanding the text, or when they find that an unknown word is occurring repeatedly and might, therefore, be worth their direct attention (Hulstijn, Hollander, & Greidanus, 1996; Keating, 2008; Kim, 2011).

A substantial body of research has been carried out regarding the importance of deliberate vocabulary learning, and it is believed that language-focused vocabulary learning can considerably speed up the process of vocabulary learning and retention. Besides, it brings about learners' awareness of the systematic features of vocabulary items (Kim, 2011; Nation & Newton, 2009; Peters, 2012) and contributes to the improvement of implicit knowledge. As a result, incidental learning of vocabulary items is believed to be accompanied by intentional learning, and both should be included in a second/foreign language teaching program (Laufer & Rozovski-Roitblat, 2011; Peters et al., 2009). Despite the call for incorporating both explicit and implicit teaching of vocabulary in any language teaching program, no consensus has ever been reached on the most appropriate and practicable methods of implementing this approach. For instance, while some pedagogical approaches which promote a deep level of processing accentuates the depth with which items are processed, they do not overlook the influence of repetition on learning.

In order for words to be learned and retained more efficiently, vocabulary instruction should engage learners in thoughtful processing. Similarly, teachers should check the procedure and techniques they use when teaching vocabulary. One possible solution to this demanding task is, perhaps, for teachers to look at the teaching-learning process from a 'levels of processing' viewpoint (Craik & Tulving, 1975), that is, defining the degree of thoughtfulness learners should have when learning vocabulary. In other words, teachers should check to see whether learners give attention to more than one aspect of the word; whether they are creative in the use of the words; and if they relate the word to their previous knowledge (Nation & Newton, 2009). In a similar way, Schmitt and McCarthy (1997) observe that giving reasons for word choices, thinking aloud, and relating words to other known vocabulary items and learner experience lead to deep processing and effective learning and retention.

The notion of 'encoding elaboration' was proposed to overcome the drawbacks of the 'depth of processing'. The main difference between these frameworks lies in the nature of encoding operations. While depth implies a constant sequence of analyses leading to learning, elaboration focuses on the quality of processing (Craik & Tulving, 1975). In addition, retention is claimed to be less correlated with the length of time spent on a piece of information held in primary memory than with the depth with which information is initially processed (Craik & Tulving, 1975). Although there is a huge body of empirical evidence dealing with the notions of depth of processing and elaboration, they have not been clearly operationalized (Laufer, 2001; Keating, 2008; Marmol & Sanchez-Lafuente, 2013).

2. Theoretical framework and literature review

In order to bridge this gap in the field of applied linguistics, Laufer and Hulstijn (2001) operationalized the notions of depth of processing (Craik & Lockhart, 1972) and elaboration (Craik & Tulving, 1975) and came up with the new framework of involvement load hypothesis (ILH), which is, perhaps, the most noticeable and widely used model of vocabulary teaching/learning. They used the term "involvement" to refer to learners' engagement in the process of learning and retention of new vocabulary items. The ILH hypothesis has a construct of two dimensions, namely cognitive and motivational, with three basic components including Need, Search, and Evaluation. Need is the motivational non-cognitive component of the construct, indicating the need for the achievement of the task. It can arise from the external world or can be self-imposed. Search and Evaluation are the cognitive dimensions of the construct. Search refers to learner effort in order to find the meaning of the word by using a dictionary or consulting any other sources. Evaluation entails a comparison of the word meaning with other possible words to decide on the correct one which best suits a particular context (Laufer & Hulstijn, 2001).

Although a large number of merits have been enumerated for the involvement load hypothesis, Zarifi (2013) and Zarifi, Jayakaran, and O'Dowd (2020) recount a number of theoretical and practical problems associated with the hypothesis. To begin with, the hypothesis is unclear in its identification of the model components, making the operatinalization and measurement of the involvement load of learning tasks rather difficult. Second, the degree of prominence related to the three components of the hypothesis is different, and they do not contribute equally to the involvement load of an activity. Third, the model fails to successfully determine the involvement load degree of different learning tasks. More specifically, the assignment of involvement load to the Need component as moderate or strong is purely subjective. Fourth, although consideration of both structural and semantic aspects of vocabulary items would bring about deeper and richer levels of processing than accounting for any of them alone (Laufer & Hulstijn, 2001), the model fails to give weight to the structural features of lexical items when assigning involvement load to different learning tasks. Therefore, borrowing the same components of the involvement load hypothesis, Zarifi (2013) developed an alternative model, namely cognitive load framework (CLF), which is claimed to be "more clearly operationalized, more conveniently practicable, and more easily measurable" (p. 1). Although CLF is mainly built upon the ILH model in both its underlying assumptions and components, it, however, provides some rather different operational definitions of the components. Zarifi (2013) presents the cognitive load framework as follows:

> As the first component of CLF framework, Need is defined as the requirement of a given word form/ meaning for completion of a given task. Therefore, need is

> considered as a purely cognitive factor. To illustrate the point, Need is moderate (1) in activities in which learners have to choose either the correct form or meaning of the word from among a number of vocabulary items such as in form-meaning matching and fill-in tasks. However, in exercises in which learners are expected to select both the appropriate form and meaning of words, Need is present and strong (2). On the other hand, in mechanical drills where the cue is available for substitution with no concern for the recognition of the form and/or meaning of the item, Need is absent (0).
>
> CLF defines Search as whether or not the target word has been provided in the task. For example, in gap-filling exercises where words are provided and learners have to choose the appropriate word or correct from among them, Search is absent (0). But Search is moderate (1) in another kind of fill-in-blanks tasks which requires learners to complete the blanks with the correct form of the given words. Finally, activities in which learners have to fill in gaps from scratch and with their own knowledge have a strong Search (2).
>
> Evaluation has been operationally defined as the use of a given word in a task that entails an evaluation of the word against its other forms or other words. Therefore, in mechanical drills where no evaluation of form or meaning of the cue is concerned, Evaluation is absent (0). Evaluation is, however, moderate (1) in multiple-choice tasks in which learners are asked to choose the antonym, synonym or the correct definition of words since such activities require learners to make a distinction between different semantic or formal aspects of the given words. Finally, Evaluation is strong (2) in activities where learners have to produce an original sentence with the word in oral or written form (pp. 117-18).

As Zarifi (2013) argues, by combining these three factors and their prominence value, one can determine the cognitive load value associated with a given learning task. The underlying assumption of this framework, like that of the involvement load hypothesis, is that the higher the cognitive load of activities, the deeper levels of processing and the more effective vocabulary acquisition and retention. This assumption is supported by Craik and Lockhart's (1975) claim that the depth of processing refers to the degree of semantic and cognitive analysis which definitely impacts the chance of the information to be stored in long-term memory. This hypothesis is also supported by Nation and Webb (2011) who maintain "the greater the cognitive load, the better the learning" (p. 12).

Although the cognitive load framework has not yet been put to any acidity test, and its predictions about the impact of task cognitive load on word learning and retention is at present a matter of conjecture, there are some pieces of empirical evidence bearing out such speculation. For example, Ellis and He (1999) found that the words that were used in productive tasks (Evaluation ++) were remembered better than the words practiced in non-productive tasks (Evaluation +). Moreover, Hulstijn, Hollander, and Greidanus (1996) showed that reading tasks requiring looking up words in a dictionary (Need + Search + Evaluation +) were more conducive to word retention than reading tasks with marginal glosses (Need + Search – Evaluation +).

Despite the fact that all these pieces of research evidence indirectly support the predictions made by the CLF, they all drew on the involvement load hypothesis as the framework of ref-

erence. The researchers were, therefore, interested in seeking further empirical evidence for the CLF within the framework described and established by the CLF model itself. More specifically, the present study was motivated to investigate whether learning activities, with higher cognitive load degrees as predicted by cognitive load framework (Zarifi, 2013), are more effective for vocabulary learning and retention than those with lower load. To put this into perspective, this study was carried out to answer the following questions:

1. Are activities with higher cognitive load degree more effective for vocabulary learning?
2. Are activities with higher cognitive load degree more effective for vocabulary retention?

3. Methodology

This section introduces the participants and sampling method, the instruments employed for data collection and the steps taken in carrying out the study.

3.1. Participants and sampling procedure

A sample of 60 Iranian English learners participated in this quasi-experimental study. The participants were chosen after Oxford Placement Test (OPT) (Allen, 2004) was administered to a population of 87 language learners. Sixty (N=60) learners with one SD above or below the mean score were chosen as the participants of the study. The participants were learning English at a language institute in Dehdasht, southwest of Iran. Of the participants, 30 learners were high school students and 30 were doing different BA programs such as law, psychology, mathematics, and engineering at university. It usually happens that learners of different ages and levels of education sit in the same class in language institutes. This is mainly because language learners at language institutes are placed at different levels based on their language knowledge rather than their levels of education or age. The participants' age range was from 16 to 22. All of them spoke Persian as their mother tongue and learned English as a foreign language. At the time of the study, the participants had attended the institute classes for about two months.

3.2. Instruments

In this study, two instruments were used to collect the data: Oxford Placement Test and a researcher-made vocabulary test. OPT was used at the outset of the study to make sure that all the participants were at the same level of language proficiency before they underwent any treatment. Allan (2004), the test developer, believes that the purpose of OPT is to measure the language knowledge that the students have and to place them as accurately and reliably as possible into different levels. This test has been calibrated against the proficiency levels based on the Common European Framework of Reference for Languages (CEFR), the Cambridge ESOL Examinations, and other major international examinations such as TOEFL.

The researcher-made vocabulary tests were of three major types: multiple-choice, fill-in, and creative writing. These test types were designed simply because they were believed to bear

different degrees of cognitive load, namely low, medium and high degrees. The validation of the tests was established by two experts in the ESL field, and the reliability of the test items was estimated by K-R20 as .76. Table 1 shows the cognitive load degree associated with each test type.

Table 1. Cognitive load degree of each test type.

Task	Status of target vocabulary	Need	Search	Evaluation
Fill-in 1	Relevant but provided in a box	2	1	1
Fill-in 2	Relevant and provided in parentheses but with unspecified form	2	1	1
Multiple-choice	Relevant and provided	1	1	1
Creative writing	Relevant but not provided	2	2	2

According to Table 1, multiple-choice items are low in terms of Need, Search and Evaluation indices as predicted by the CLF. That is, learners are provided with some choices among which they should choose the best form or meaning of the words. To put this into perspective, either the form or meaning of the vocabulary is required for the completion of the sentence. Learners have to search only for the correct form or meaning, not both, to complete the sentence, or they have to evaluate either one form of a word with its other forms or its meaning with that of the other options. Fill-in activities which are moderate in terms of cognitive load consist of two types of activities. In the first type, there are some sentences each with a blank preceded by some words. Learners should read the sentences and complete them with the words provided. So, the Search index is weak as the words that complete the sentences are provided. In addition, since both the meaning and appropriate form of the given words are necessary for the completion of the exercise, Need is strong. The Evaluation is, however, weak simply because the word is not evaluated against any other word. In the second type, learners are given some sentences with a vocabulary item embedded in parentheses and asked to fill in the blank with the appropriate form of the word. Here they should search for the correct form though not for the correct word in terms of meaning; therefore, Search is weak because learners have to look only for the correct form of the word. However, Need and Evaluation are strong and weak respectively as in the first type. In creative writing Need, Search, and Evaluation are strong since learners have to search for the appropriate words that fit in the context in terms of both meaning and form. They should also evaluate the correct form and meaning of the words against other possible vocabulary items, hence strong Evaluation.

3.3. Procedure of the Study

All the teaching materials used for the three groups were the same except for the vocabulary tasks which were of different cognitive loads as required by the research questions. The researchers designed the activities and gave them to an experienced teacher to work with the students. The teacher was 35 years old and had already taught English in different institutes and non-state universities for 10 years. He was also a Ph.D. student of TEFL. Therefore, the students were exposed to and performed vocabulary tasks with different degrees of cognitive load. That

is to say, experimental group 1 (EG1) received vocabulary tasks which had the highest cognitive load; experimental group 2 (EG2) was exposed to vocabulary tasks having a medium cognitive load; and the third experimental group (EG3) experienced the vocabulary tasks with the lowest cognitive load. It should, however, be pointed out that the researcher provided the teacher with essential guidelines and necessary explanations for implementing each vocabulary task. The vocabulary tasks were created based on the content of *Top Notch 2A, 2B* (Saslow & Ascher, 2011) used with the students at the institute as the instructional materials. Some of the target vocabulary items which the learners were exposed to, included 'documentary, kindness, attractive, modest, emotion, and the like'.

Three groups of tasks were developed, namely creative writing, fill-in, and multiple-choice, with creative tasks being the highest, fill-in tasks being the moderate and multiple-choice tasks being the lowest in terms of cognitive load. After the presentation of the target words, the EG1 learners were asked to use the target vocabulary items in some new sentences. This task was thought to have the highest cognitive load since the learners were involved in some creative tasks. The fill-in tasks, with the moderate cognitive index, constituted some sentences with missing parts and the learners in EG2 were required to fill the sentences with provided words in each session. Important to note is that the learners were provided with alternatives to choose. For instance, there were ten sentences each with a blank to be filled in, and they were accompanied by twelve words (two extra words). The learners in EG2 read and completed the sentences using the words provided above the sentences. In the other type of fill-in activities, some sentences were given to the students with vocabulary items embedded in parentheses and they were supposed to fill in the blanks with the appropriate forms of the words. Here they had to search for the correct form though not for the correct word in terms of meaning. For the third experimental group (EG3), multiple-choice exercises were employed as these learning activities were considered to have the lowest cognitive load. In this task, the learners were asked to read some sentences and complete them with the alternatives provided.

4. Results and discussion

Before reporting the findings related to each research question, it is important to provide a brief account of the participants' performance in the pre-test across the three groups. The descriptive statistics of the pre-test revealed that experimental group 1 (EG1) had the highest mean (M=15.700, SD=2.903), followed by the third experimental group (EG3) (M=15.550, SD=3.284) and then the second experimental group (EG2) (M=15.400, SD=2.779).

In order to find out whether the observed difference among the mean scores of the three groups on the pre-test was of any significance, One-way ANOVA was run. As shown in Table 2, the results indicated that there was no significant difference between the means of the three groups ($F(2) = .05$, $p=.951$) before the treatment. In other words, the three groups were on the same footing in terms of their knowledge of vocabulary at the outset of the study. Therefore, any difference in their performance on the post-test and delayed post-test could be mainly attributed to the effect of the learning tasks designed and offered in line with the predictions of the cognitive load framework.

Table 2. One-way ANOVA for the pre-test.

	Sum of squares	Df	Mean square	F	Sig.
Between groups	.900	2	.450	.050	.951
Within groups	511.950	57	8.982		
Total	512.850	59			

4.1.Research question one

In the fill-in task, the Need index was moderate as only the recognition of the meanings of the words not their forms was relevant; the Search index was also moderate since the participants were required to fill the blank spaces with the words provided; similarly, the Evaluation index was moderate because the participants had to evaluate the cues semantically in order to select the word that best suited the sentence.

The descriptive statistics of the three groups in the post-test revealed that both experimental groups 1 and 2 who were exposed to the tasks of high and medium cognitive load outperformed the third experimental group (M=17.200, SD=2.764) that was exposed only to multiple choice items. Experimental group 1, however, had the highest mean score (M=26.450, SD=1.700).

In order to determine whether the observed difference among the mean scores of the participants in the three groups was statistically significant, One-way ANOVA was used. Table 3 reports the results of one-way ANOVA.

Table 3. One-way ANOVA for the post-test of vocabulary learning.

	Sum of squares	Df	Mean square	F	Sig.
Between groups	855.7	2	427.8	81.865	.001
Within groups	297.9	57	5.226		
Total	1156.6	59			

The results demonstrate that the mean scores of EG1, EG2, and EG3 were significantly different from one another (F (2) = 81.865, p= .001). Therefore, to determine which group(s) in the sample differed from the other ones, the analysis was complemented with the Tukey's HSD test.

Table 4. Post hoc Analyses (Tukey's HSD) for Vocabulary Learning.

(I) Participants	(J) Participants	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower bound	Upper bound
EG1	EG2	4.700*	.722	.001	2.96	6.43
	EG3	9.250*	.722	.001	7.51	10.98
EG2	EG1	-4.700*	.722	.001	-6.43	-2.96
	EG3	4.550*	.722	.001	2.81	6.28
EG3	EG1	-9.250*	.722	.001	-10.98	-7.51
	EG2	-4.550*	.722	.001	-6.28	-2.81

* The mean difference is significant at the 0.05 level.

As shown in Table 4, the Tukey HSD test indicates that the mean score for EG1 (M=26.450, SD=1.700) was significantly different from those of EG2 (M=21.750, SD=2.268) and EG3 (M=17.200, SD=2.764). The results indicate that the vocabulary tasks with higher cognitive load resulted in more effective vocabulary learning. More specifically, the participants in EG1 who received vocabulary activities with the highest cognitive load outperformed the other two groups in terms of vocabulary learning. Additionally, the results indicate that the fill-in tasks which were used with EG2 were significantly more effective for vocabulary learning than the multiple-choice tasks which were offered to EG3.

4.1. Research question two

In addition to investigating the effect of the tasks with different cognitive load indices on vocabulary learning, the study also explored the effectiveness of the learning activities with different cognitive load degrees on vocabulary retention. To this end, the participants of the three groups received a delayed post-test after four weeks. The descriptive statistics of the delayed post-test for EG1, EG2, and EG3 reveal that the participants in EG1 (M=26.050, SD=2.18) outperformed the participants in EG2 (M=20.900, SD=2.77) and EG3 (M=16.500, SD=3.37). Therefore, a one-way ANOVA test was also run to find out if the observed difference among the mean scores of the three groups was statistically significant. The result of one-way ANOVA is presented in Table 5.

Table 5. One-way ANOVA for the delayed post-test.

	Sum of squares	Df	Mean square	F	Sig.
Between groups	913.900	2	456.950	57.402	.001
Within groups	453.750	57	7.961		
Total	1367.650	59			

As Table 5 reveals, the difference between the retention rates of the three groups proved to be statistically significant; therefore, the one-way ANOVA test was complemented with the post-hoc analysis test of Tukey's HSD for the identification of the significant inter-group differences.

Table 6 presents the results for Tukey's HSD.

Table 6. *Post hoc analyses (Tukey's HSD) for vocabulary retention.*

(I) Participants	(J) Participants	Mean difference (I-J)	Std. error	Sig.	95% Confidence interval	
					Lower bound	Upper bound
Exp1	Exp2	5.150*	.892	.001	3.00	7.29
	Exp3	9.550*	.892	.001	7.40	11.69
Exp2	Exp1	-5.150*	.892	.001	-7.29	-3.00
	Exp3	4.400*	.892	.001	2.25	6.54
Exp3	Exp1	-9.550*	.892	.001	-11.69	-7.40
	Exp2	-4.400*	.892	.001	-6.54	-2.25

* The mean difference is significant at the 0.05 level.

The post hoc comparisons indicate that the mean score for EG1 that was exposed to the creative writing task with the highest cognitive load index (M=26.050, SD=2.187) was significantly different from those of EG 2 that received the fill-in tasks with moderate cognitive load (M=20.900, SD=2.770) and EG 3 (M=16.500, SD=3.379) that was exposed to the multiple-choice task. Taken together, the results suggested that the tasks and activities with high levels of cognitive load did have a significant effect on vocabulary learning and retention.

5. Discussion

The objective of the present study was to put into practice and test the conceptual model of cognitive load framework proposed by Zarifi (2013). The assumption underlying CLF is that activities in which learners are more involved and which have higher degrees of cognitive load lead to deeper mental processing and are more effective for learning and retention. The results obtained from the present study reveal that the amount of mental effort required to complete vocabulary tasks affects vocabulary learning and retention differently. In particular, it was found that the vocabulary activities with the highest cognitive load indices were the most effective. In other words, the participants in experimental group 1 who were engaged in the creative writing task with the highest cognitive load performed better in vocabulary learning and retention than those in experimental group 2 who were exposed to the fill-in task with the medium cognitive load and those in experimental group 3 who were involved in the multiple-choice exercises with low cognitive load. Therefore, the results rendered empirical support for the conceptual model of CLF framework developed by Zarifi (2013).

The bottom line of the cognitive load framework (Zarifi, 2013) is that vocabulary learning and retention is contingent upon whether processing of word form and/or word meaning is required for doing a given task or not (Need), whether or not the word to be used in the task is provided (Search), and whether the use of a given word in a task entails an evaluation of the word against its other forms or other words (Evaluation). In line with this prediction, the par-

ticipants who were engaged in creative writing in experimental group 1 outperformed the other two groups. The creative writing task entailed strong Need, Search, and Evaluation. As a matter of fact, these learners had to find words to express their ideas in writing; they were forced to look up the unknown words in a dictionary or to ask someone; they had to compare the words to come up with an appropriate choice; and they also needed to consider the usage of the words they intended to use in their writing, hence more cognitively involved in performing the tasks.

It is interesting to point out that the results of the present study are in good agreement with the findings of some other researchers following the involvement load hypothesis as their model of reference (e.g., Cho & Ma, 2013; Huang, Willson, & Eslami, 2012; Keating, 2008; Kim, 2008; Nassaji & Hu, 2012; Teng, 2015; Zou, 2012). For instance, the Keating (2008) study of the learning and retention of vocabulary among a student population of beginning learners of the Spanish language experiencing three different tasks with varying degrees of involvement load also renders support for the predictions made by the CLF framework and the empirical evidence provided by the current study. More specifically, he found that the rate of vocabulary learning and retention was highest in the sentence writing task which required strong cognitive involvement, lower in the reading plus fill-in task which had moderate involvement, and lowest in the reading comprehension task only. In a similar study, Kim (2008) found "a higher level of learner involvement during the task promoted more effective initial vocabulary learning and better retention of the new words" (p. 285). Moreover, he revealed that different tasks with the same involvement load index resulted in similar amount of initial vocabulary learning and retention. Likewise, in a meta-analysis study, Huang, Willson, and Eslami (2012, p. 554) argued "language learners who performed a task with a higher degree of involvement load, gained more vocabulary". Finally, Nassaji and Hu (2012), in a correlational research on three groups of learners doing three reading texts which were expected to require different degrees of involvement load values showed that the text with high involvement load "led to the use of significantly more word-based strategies than the other two texts. The high involvement load text also led to significantly higher rates of retention of word meanings that were inferred correctly" (Nassaji & Hu, 2012, p. 82).

The findings of the current study are, however, in conflict with those of some other studies (e.g., Haratmeh, 2012; Martínez-Fernández, 2008; Xudong, 2010; Yaqubi, Rayati, & Allemzade, 2012). To begin with, Martínez-Fernández's (2008) study failed to provide empirical support for involvement load hypothesis. He argued that output rather than the degree of involvement load is responsible for higher vocabulary learning and retention. In addition, input-orientation tasks, as compared with output-orientation tasks, did not prove to have more positive effect on vocabulary development. Similarly, Yaqubi, Rayati, and Allemzade (2012) found no empirical support for involvement load hypothesis. More specifically, they did not find any significant superiority of tasks with higher involvement load index. Similarly, they reported that the participants who were involved in the output oriented task outperformed those completed the input-oriented task. Finally, the findings of an empirical study by Haratmeh (2012) were contrary to the predictions of the involvement load hypothesis. The researcher held "involvement load is not the only determining factor in task effectiveness, but input/output-orientation of tasks is also a decisive parameter in task effectiveness" (Haratmeh, 2012, p. 86).

The failure of the above studies for providing evidence for the involvement load hypothesis may, in one way or another, be attributed to the inappropriate operationalization of the components of the model presented by Laufer and Hulstijn (2001). As it was pointed, the three components do not equally contribute to the involvement load of the learning activities. More specifically, the prominence degree of the three components of the hypothesis is different, with Need carrying a load index of either + or ++, Search ranging in load value of - or +, and Evaluation varying in

degree from - to + or ++. Moreover, the job of assignment of involvement load index to each task following the guidelines set by the involvment load hypothesis is purely subjective (Zarifi, Jayakaran & O'Dowd, 2020). To exemplify the point, the researcher is not clear as to how to decide whether a task is learner imposed (Strong Need) or teacher imposed (Weak Need).

The results obtained from the current study provided new insights into the area of vocabulary teaching and learning. The study revealed that the output and input dichotomy cannot be reliable criteria to categorize vocabulary tasks in terms of their effectiveness. While some researchers (Martínez-Fernández, 2008; Sarani, Negari, & Ghaviniat, 2013) pointed out the output tasks lead to more vocabulary learning and retention, the present study found that different output tasks can lead to different results as far as their cognitive load index is concerned. That is to say, the degree of cognitive load that a vocabulary task entails can be an evidence-based index to determine its effectiveness for vocabulary learning and retention. In the present study, the creative writing and the fill-in tasks were both considered to be the output tasks through which the participants were to produce language. However, these two types of tasks did not render the same results, with the creative writing being more effective for vocabulary learning and retention.

6. Conclusion

The findings of this empirical study suggest that vocabulary learning tasks and activities can be designed with regard to the cognitive load they carry and the effect they might have on vocabulary learning and retention. The study revealed that the creative writing tasks in which the indices of Need, Search and Evaluation were strong, hence with high cognitive load, resulted in better vocabulary learning and retention than the fill-in tasks with moderate cognitive load index (Moderate Need, Search and Evaluation), and the multiple-choice exercises with low cognitive load (Low Need, Search and Evaluation). To put this into perspective, experimental group 1 who completed the creative writing task with strong Need, Search and Evaluation indices outperformed experimental group 2 who were exposed to the fill-in tasks with moderate Need, Search and Evaluation indices and experimental group 3 dealing with multiple-choice exercises of low Need, Search and Evaluation degrees. More specifically, vocabulary tasks which require a selection as to the appropriate use of both meaning and form of words, those in which learners are not provided with vocabulary items or alternatives to use, and those in which use of the target words involves generating an original sentence in verbal or written form contribute to more effective vocabulary learning and retention. Taken together, the findings of the study provided empirical support for the cognitive load framework (Zarifi, 2013), which posits that learning and retention of new vocabulary ítems would be an outcome of the combination of the need for, the evaluation of, and the search for their form and meaning against their other posible forms and meanings or those of other words. To make a long story short, the more demanding the vocabulary task is in terms of the cognitive load it carries, the more effective it tends to be for vocabulary learning and retention.

7. Pedagogical implications

This study has implications for language teachers, learners, and materials developers. As far as language teachers and learners are concerned, the study revealed that vocabulary learning and

retention is remarkably influenced by the amount of cognitive load they require from language learners. Teachers can take advantage of the findings of this study and make the demanding and challenging journey of vocabulary learning more efficient. For instance, they can ask students to do creative writing tasks, using the words to which they have already been exposed. By designing exercises or learning activities of high load cognitive index, teachers can provide learners with enough opportunity to cope with the task of vocabulary learning. The results of the study can also help materials developers in some other ways. They can design different vocabulary tasks while taking into account the degree of cognitive load of each. They should provide for the inclusion in textbooks of more creative writing and fill-in tasks rather than only multiple-choice ones which are of low cognitive load.

8. References

Allan, D. (2004). *Oxford placement test 2.* Oxford University Press.

Cho, Y. A., & Ma, J. H. (2013). The effects of task-induced involvement and word exposure frequency on L2 college students' vocabulary acquisition. *Studies in English Language & Literature*, *38*(1), 339-367.

Craik, F. M. I. & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*, 671-684.

Craik, F. M. I. & Tulving, E. (1975). Depth of processing and the retention words in episodic memory. *Journal of Experimental Psychology*, *10*(4), 268-294.

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

Haratmeh, M. S. (2012). Involvement load and task type in task effectiveness: two aspects of vocabulary knowledge. *International Journal of Academic Research*, *4*(4), 86-95.

Huang, S., Eslami, Z.R. & Willson, V. L. (2012). The effects of task involvement load on L2 incidental vocabulary learning: a meta-analytic study. *Modern Language Journal 96(4),* 544-557.

Hulstijn, J. H. (1992). Retention of inferred and given word meanings: Experiments in incidental vocabulary learning. In P. J. Arnaud & H. B'joint (Eds.), *Vocabulary and applied linguistics*, (113-125). Macmillan.

Hulstijn, J. H., Hollander, M., & Greidanus, T. (1996). Incidental vocabulary learning by advanced foreign language students: The influence of marginal glosses, dictionary use, and reoccurrence of unknown words. *Modern Language Journal, 80, 327-339.*

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

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.

Kim, Y. (2011). The role of task-induced involvement and learner proficiency in L2. *Language Learning, 61*(1), 100-140.

Krashen, S. (2008). Language education: Past, present and future. *RELC Journal, 39*, 178–87.

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

Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, *16*, 307–322.

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

Lewis, M. (1993). *The lexical approach: The state of ELT and the way forward.* Hove: Language Teaching Publications.

Mármol, G. A., & Sánchez-Lafuente, Á. A. (2013). The Involvement Load Hypothesis: Its effect on vocabulary learning in primary education. *Revista Española De Lingüística Aplicada, 26*, 11-24.

Martínez-Fernández, A. (2008). Revisiting the involvement load hypothesis: Awareness, type of task and type of item. In *Selected proceedings of the 2007 second language research forum* (pp. 210-228). Somerville, MA: Cascadilla Press.

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

Nation, P. & Newton, J. (1997). Teaching vocabulary. In J. Coady and T. Huckin (eds.), *Second language vocabulary acquisition*. Cambridge: Cambridge University Press.

Nation, P. & Macalister, J. (2010). *Language curriculum design.* New York: Routledge.

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

Peters, E. (2012). The differential effects of two vocabulary instruction methods on EFL word learning: a study into task effectiveness. *International Review of Applied Linguistics in Language Teaching, 50*(3), 213-238.

Peters, E., Hulstijn, J.H., Sercu, L., & Lutjeharms, M. (2009). Learning L2 German vocabulary through reading: The effect of three enhancement techniques compared. *Language Learning, 59,* 113–51.

Pellicer-Sánchez, A. & Schmitt, N. (2010). Incidental vocabulary acquisition from an authentic novel: Do Things fall apart? *Reading in a Foreign Language*, *22*, 31–55.

Richards, J. C., & Renandya, W. A. (2002). *Methodology in language teaching.* Cambridge University Press.

Sarani, A., Negari, G. M., & Ghaviniat, M. (2013). The role of task type in L2 vocabulary acquisition: a case of Involvement Load Hypothesis. *Language and Culture*, *35*(4), 377-386.

Saslow, J.M. & Ascher, A. (2011). *Top notch.* Pearson Education ESL.

Schmitt, N., & McCarthy, M. (Eds.). (1997). Vocabulary: description, acquisition and pedagogy. Cambridge University Press.

Stahl, S. A., & Nagy, W. E. (2006). *Teaching word meanings.* London: Lawrence Erlbaum Associates.

Teng, F. (2015). Involvement load in translation tasks and EFL vocabulary learning. *The New English Teacher*, *9*(1), 83-101.

Xudong, W. U. (2010). Can learning tasks affect incidental vocabulary acquisition? involvement load hypothesis revisited. *Foreign Language Teaching and Research*, 4*2 (2)*, 109-116.

Yaqubi, B., Rayati, R. A., & Allemzade, 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.

Zarifi, A. (2013). *Establishing and evaluating phrasal verb use in a Malaysian ESL secondary school textbook* (Doctoral dissertation). Universiti Putra Malaysia, Serdang.

Zarifi, A., Jayakaran, M., & O'Dowd, E. (2020). Cognitive Load Framework: an alternative to the Involvement Load Hypothesis. Khazar Journal of Humanities and Social Sciences (in Press).

Zou, D. (2012). *A study of the components of the involvement load hypothesis: How involvement load should be allocated to" Search" and" Evaluation"* (Unpublished doctoral dissertation) City University of Hong Kong, Hong Kong.